

**Best  
Available  
Copy**

**AD-A284 472**



*Dist: A*

*D*

N PAGE		Form Approved OMB No. 0704-0188	
<small>POLICY FOR REDUCTION OF BURDEN. INCLUDING THE TIME FOR REVIEWING INSTRUCTIONS, SEARCHING EXISTING DATA SOURCES, ACTION OR INFORMATION. SEND COMMENTS REGARDING THIS BURDEN ESTIMATE OR ANY OTHER ASPECT OF THIS FORM TO: HEADQUARTERS SERVICES, DIRECTORATE FOR INFORMATION OPERATIONS AND REPORTS, 1215 JEFFERSON MEET AND BUDGET, FEDERALWORK REDUCTION PROJECT (0704-0188), WASHINGTON, DC 20503.</small>			
1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED	6/93-6/94
		Final Technical Report	
4. TITLE AND SUBTITLE		5. FUNDING NUMBERS	
Computational Vision Based on Neurobiology		Q315-AS F49620-93-1-0274 N00014-93-1-0612	
6. AUTHOR(S)		6. PERFORMING ORGANIZATION REPORT NUMBER	
Teri B. Lawton, editor		61102F	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)		8. PERFORMING ORGANIZATION REPORT NUMBER	
The International Society for Optical Engineering P.O. Box 10 Bellingham, WA 98227-0010		Volume 2054 AFOSR-TR-94-0523	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
Air Force Office of Scientific Research / NL Bolling AFB, D.C. Dr Tangney			
Office of Naval Research Arlington, VA			
11. SUPPLEMENTARY NOTES			
Papers published in the attached proceedings report were presented at a conference held 6-9 July 1993 at Pacific Grove, California.			
12a. DISTRIBUTION/AVAILABILITY STATEMENT		12b. DISTRIBUTION CODE	
unclassified/unlimited		S DTIC SELECTED SEP 07 1994 D F	
13. ABSTRACT (Maximum 200 words)			
Biological systems use multiple object attributes to construct a 3D perception from an initial 2D representation. This report explores computational vision models that are based on neurobiology. Each of the fundamental levels of analysis needed for high level pattern recognition are addressed to provide new insights into the different processing modules. Papers detail methods for reconstructing 3D images from partial information, for correcting image defects, or for effectively extracting/analyzing/interpreting images of neurobiological and biomedical interest.			
<small>The document has been approved for public release and sale; its distribution is unlimited</small>			
14. SUBJECT TERMS		15. NUMBER OF PAGES 246	
computational vision, neurobiology, image processing		16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	

ISBN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)  
Prescribed by ANSI Std Z39-18  
298-102

# PROCEEDINGS



SPIE—The International Society for Optical Engineering

## ***Computational Vision Based on Neurobiology***

**Teri B. Lawton**  
*Chair/Editor*

**6–9 July 1993**  
**Pacific Grove, California**

254P  
**94-29026**



DIGITAL QUANTUM INFORMATION



**Volume 2054**

# PROCEEDINGS



SPIE—The International Society for Optical Engineering

## *Computational Vision Based on Neurobiology*

**Teri B. Lawton**  
*Chair/Editor*

**6–9 July 1993**  
**Pacific Grove, California**

*Sponsored and Published by*  
**SPIE—The International Society for Optical Engineering**

Accession For	
NTIS	CRA&I
DTIC	TAB
Unannounced	
Justification	
By _____	
Distribution	
Availability	
Dist	Aval. _____ Special
A-1	



**Volume 2054**

**DTIC QUALITY INSPECTION 3**

SPIE (The Society of Photo-Optical Instrumentation Engineers) is a nonprofit society dedicated to the advancement of optical and optoelectronic applied science and technology.



The papers appearing in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and are published as presented and without change, in the interests of timely dissemination. Their inclusion in this publication does not necessarily constitute endorsement by the editors or by SPIE.

Please use the following format to cite material from this book:

Author(s), "Title of paper," in *Computational Vision Based on Neurobiology*, Teri B. Lawton, Editor, Proc. SPIE 2054, page numbers (1994).

Library of Congress Catalog Card No. 93-86443  
ISBN 0-8194-1315-1

Published by  
**SPIE—The International Society for Optical Engineering**  
P.O. Box 10, Bellingham, Washington 98227-0010 USA  
Telephone 206/676-3290 (Pacific Time) • Fax 206/647-1445

Copyright ©1994, The Society of Photo-Optical Instrumentation Engineers.

This work relates to Department of Navy Grant N00014-93-1-0612 issued by the Office of Naval Research and to Department of Air Force Grant F49620-93-1-0274 issued by Air Force Office of Scientific Research. The United States Government has a royalty-free license throughout the world in all copyrightable material contained in the publication.

Copying of material in this book for internal or personal use, or for the internal or personal use of specific clients, beyond the fair use provisions granted by the U.S. Copyright Law is authorized by SPIE subject to payment of copying fees. The Transactional Reporting Service base fee for this volume is \$6.00 per article (or portion thereof), which should be paid directly to the Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923. Other copying for republication, resale, advertising or promotion, or any form of systematic or multiple reproduction of any material in this book is prohibited except with permission in writing from the publisher. The CCC fee code is 0-8194-1315-1/94/\$6.00.

Printed in the United States of America.

## Contents

v      **Conference Committee**  
vii     *Introduction*

1      **Coding of inhibition in visual cortical spike streams [2054-01]**  
A. B. Bonds, Vanderbilt Univ.

12     **Visual cortex neurons in monkey and cat: contrast response nonlinearities and stimulus selectivity [2054-03]**  
D. G. Albrecht, W. S. Geisler, Univ. of Texas/Austin

32     **New model of human luminance pattern vision mechanisms: analysis of the effects of pattern orientation, spatial phase, and temporal frequency [2054-05]**  
J. M. Foley, G. M. Boynton, Univ. of California/Santa Barbara

43     **Monocular and binocular mechanisms of contrast gain control [2054-04]**  
I. Ohzawa, R. D. Freeman, Univ. of California/Berkeley School of Optometry

52     **Computational reconstruction of the mechanisms of human stereopsis [2054-25]**  
C. W. Tyler, L. Barghout, L. L. Kontsevich, Smith-Kettlewell Eye Research Institute

69     **Reconstruction of 3D structure from two central projections by a fully linear algorithm [2054-26]**  
M. L. Kontsevich, Max-Plank Institut für Mathematik (FRG); L. L. Kontsevich, Smith-Kettlewell Eye Research Institute

79     **Disparity tuning of cyclopean visual mechanisms [2054-27]**  
S. B. Stevenson, C. M. Schor, L. K. Cormack, Univ. of California/Berkeley School of Optometry and Univ. of Texas/Austin

89     **Standard model of color vision: problems and an alternative [2054-30]**  
R. L. De Valois, Univ. of California/Berkeley

95     **Spatial vision based upon color differences [2054-29]**  
K. K. De Valois, Univ. of California/Berkeley

104    **Neurobiological mechanisms of cortical direction selectivity [2054-21]**  
C. L. Baker, Jr., McGill Vision Research Unit (Canada); J. C. Boulton, Utrecht Biophysics Research Institute (Netherlands)

124    **Psychophysical evidence for both a "quasi-linear" and a "nonlinear" mechanism for the detection of motion [2054-22]**  
J. C. Boulton, Utrecht Biophysics Research Institute (Netherlands); C. L. Baker, Jr., McGill Vision Research Unit (Canada)

134    **Motion mechanisms based on nonlinear spatial filters have lower temporal resolution [2054-23]**  
A. M. Derrington, Univ. of Newcastle Upon Tyne Medical School (UK)

142 **Fidelity metrics and the test-pedestal approach to spatial vision [2054-24]**  
S. A. Klein, Univ. of California/Berkeley School of Optometry

155 **Neural processing of biological motion in the macaque temporal cortex [2054-18]**  
M. W. Oram, D. I. Perrett, Univ. of St. Andrews (UK)

166 **Neuronal mechanism for signaling the direction of self-motion [2054-16]**  
J.-P. Roy, Montreal Neurological Institute/McGill Univ. (Canada)

174 **Topography of excitatory and inhibitory connectional anatomy in monkey visual cortex [2054-10]**  
J. S. Lund, J. B. Levitt, Q. Wu, Institute of Ophthalmology/Univ. of London (UK)

185 **Large-scale organization of the primate cortical visual system [2054-08]**  
M. P. Young, Univ. of Oxford (UK)

194 **Dynamic object-based 3D scene analysis using multiple cues [2054-06]**  
T. B. Lawton, Nano Tech Services

211 **Influence of figural interpretation on the selective integration of visual motion signals [2054-13]**  
T. D. Albright, G. R. Stoner, The Salk Institute for Biological Studies

220 **Workshop on computational vision based on neurobiology**  
Moderated by M. P. Young, Univ. of Oxford (UK); and T. B. Lawton, Nano Tech Services

245 *Addendum*

246 *Author Index*

## **Conference Committee**

### *Conference Chair*

**Teri B. Lawton**, Nano Tech Services

### *Program Committee*

**A. B. Bonds**, Vanderbilt University

**Ralph Siegel**, Rutgers University

**Christopher W. Tyler**, Smith-Kettlewell Eye Research Institute

**Karen K. De Valois**, University of California/Berkeley

## INTRODUCTION

### COMPUTATIONAL VISION BASED ON NEUROBIOLOGY

**Volume 2054**

Neurobiology is providing new insights into the mechanisms used for pattern discrimination and recognition. Biological systems use multiple object attributes to construct a 3-D perception from an initial 2-D representation. We have discovered that the visual system is modular and there are competitive and cooperative neural networks between different types of neurons within the same layers of the visual cortex, as reported by Jenny Lund. In addition to being modular, it is arranged hierarchically with competitive and cooperative neural networks between different areas in the brain that are used to construct a 3-D image of the world around us, as described by Malcolm Young. This conference explored the neurobiology needed to develop robust computational vision models. Teri Lawton showed how a computational visual system based on neurobiology could construct 3-D object maps two to three orders of magnitude faster and more robustly than others, who base their computational visual systems on a static pixel-based, instead of a dynamic object-based representation. Both the pattern recognition and the eye-head movement systems are modeled by Lawton, an essential component for robust 3-D reconstruction of objects in natural scenes.

The conference contained 23 papers organized into six sections covering many of the fundamental levels of analysis needed for high level pattern recognition. The conference began with a session devoted to different approaches to The Nature of Representation of Images in the Brain. Sessions to analyze the different modules used to construct a 3-D perception include: Mechanisms Used for Stereopsis, Cortical Color Mechanisms, and Dynamic Gain Control of Movement: Low Level and High Level Mechanisms. The connectional and functional organization of these different modules were described in the session Dynamic Object-Based Scene Analysis Using Multiple Attributes. This symposium was designed to bring together scientists who use a multiattribute approach for analyzing an observer's perception to provide new insights into the different processing modules, and those who are developing computational models that must analyze complex scenes at real-time frame rates. It is hoped that the broad range of relevant topics being presented at this symposium will serve to encourage interactions among scientists in psychophysics, neurobiology, computational vision, image processing, and biomedical engineering.

One aspect of cortical neurobiology that was emphasized by many speakers was the division into two parallel streams in the visual cortex. In the cortex, there are two major pathways that go beyond the magnocellular-parvocellular (also known as the transient-sustained) dichotomy found in the retina and lateral geniculate nucleus. The ventral stream conveys what an object is, being composed of both parvo and magno neurons, whereas the dorsal stream conveys where an object is located in 3-D space, being composed predominantly of magno neurons. There is only limited crosstalk between ventral and dorsal streams until they reconverge at the highest levels in the cortical hierarchy, such as the superior temporal polysensory area, using feedforward connections, and using feedback connections via cortical and subcortical pathways in the midbrain. The paper by Malcolm Young described the complex connectional organization of the visual representation areas in the cortex.

The types of mechanisms used at higher levels of analysis are not obvious, and only by mapping out the classic receptive fields, and the extended regions that are uncovered only by inhibitory and disinhibitory interactions with stimuli that activate neurons within their classic receptive field, can we understand the nature of biological mechanisms that have evolved to rapidly locate and identify objects in the world around us. Papers at this meeting elucidated stages that were not obvious, yet once uncovered explain perceptual phenomenon that could not be accounted for previously. For example, the paper by Russell De Valois elucidates an additional stage in the opponent color system, deduced from physiological recordings, that enables fine differential color discrimination to have evolved in macaques. Jenny Lund, based on anatomical studies, laid out the intrinsic circuitry of superficial layers in the visual cortex, providing a means by which the interaction between the inhibitory zones of

pyramidal and basket neurons predicts the silent inhibitory and disinhibitory zones that surround the classic receptive fields. The importance of both global and local inhibitory interactions to account for contrast gain control was emphasized by each one of the speakers in the first session.

Psychophysical, physiological, and anatomical studies presented at this symposium elucidated stages in visual mechanisms that if incorporated into computational visual systems would increase their robustness by improving the competitive and cooperative algorithms at little or no cost. Not only does computational vision need models based on neurobiology, but neurobiology needs computational models to understand the complex predictions that are obtained when: 1) several levels of analysis are used for a task, where adjacent areas interact through feedback and feedforward connections, or 2) within a single area, where the excitatory and inhibitory interactions between different types of neurons provide a continuum of gradient-based connectional lattices for both broadband and narrowband mechanisms, as pointed out by Jenny Lund during the workshop Wednesday night. Christopher Tyler illustrated the degree to which computation can elucidate the underlying mechanisms whose properties are not obvious from the data used to derive them.

Future symposia will expand upon the current one as more computational visual systems based on neurobiology are developed. A much larger proportion of submitted presentations, instead of using only invited presentations, will be employed for future meetings. More participation from students will be encouraged. The success of this conference is the result of the participation by leading neurobiologists discussing problems they've encountered that are useful for computational vision, by inquisitive students in this area, and by scientists from related areas that are looking for ways to improve the robustness of computational vision models so that they are more useful.

### **Opening Wider The Doors of Computational Vision**

Dr. Vernon Dobson in his seminal and influential book *Models of The Visual Cortex* edited with Dr. David Rose provided one of the first collections of articles that addressed computational vision models based on neurobiological constructs, and not the engineering constructs used by others who have developed neural network models. Unfortunately, Vernon Dobson, who was to be a key speaker at this meeting, tragically drowned while swimming just after presenting his research at the Fifteenth European Conference on Visual Perception in Pisa, Italy last year.

Vernon Dobson was one of those rare scientists able to bridge the realms of psychology, philosophy, biology, and engineering. The senior research fellow in experimental psychology at Oxford, holder of international patents in microchip design, author of numerous scientific papers on visual perception, and the neural mechanisms of learning, and was due to take up a senior post at the new University of Hertfordshire. After obtaining degrees in psychology and cybernetics from Bristol and Brunel Universities respectively, he went on to develop theories of visual perceptual mechanisms based on neural networks. He was careful to make his models as biologically plausible as possible, basing them on anatomical and neurophysiological evidence for how the mammalian visual system actually functions. His insights into the significance of physiological inhibition for brain function led to the use of inverse or negative logic for those neural networks. The same principle was also incorporated into the design of a microchip for artificial visual systems that are capable of being trained to perceive their environment. Vernon argued that real, biological neural nets must have robust design principles which are selected for by Darwinian evolution, and thus do not require an elaborate preprogramming of parameters, and that effective artificial nets should share the same principles. Vernon did not just try to model one small part of the brain in isolation; he knew that vision is a complex system. His models included comprehensive global diagrams of the system, as well as detailed descriptions of the neural nets at work in each functional center. He recently discovered new, simple principles for the recognition of visual shapes and letters. In addition, he was very aware of the importance of metaphysical presuppositions about the brain. The absence of his cheerful and enthusiastic personality, his willingness to communicate, challenge, and encourage, and his significant innovative approaches to perception is a great loss. We dedicate this meeting to his memory.

## Coding of inhibition in visual cortical spike streams

A. B. Bonds

Dept. of Electrical Engineering, Vanderbilt University  
P.O. Box 1824B, Nashville, TN 37235

### ABSTRACT

Modulation of cortical firing rate is a major factor in defining cortical filter properties. Active response suppression (inhibition) is seen whenever cortical cells are exposed to grating stimuli that are non-optimal, in either the domain of orientation or spatial frequency. Responses are also reduced by pre-exposure to gratings of high contrast. The first phenomenon is termed spatially-dependent inhibition, the second contrast gain control. We have explored the physiological basis for these two phenomena in striate cortical cells of the anesthetized cat. Sequences of spikes in responses show bursts characterized by interspike intervals of 8 msec or less. Both burst frequency and burst length depend on average firing rate, but at a given firing rate burst length is lower for non-optimal orientations. Burst length is also shortened by local injection of GABA. Burst length modulation is not seen in the case of contrast gain control. These results support the existence of two independent mechanisms for modulating cortical responsiveness. A GABA-ergic mechanism that shortens spike bursts is invoked by presentation of spatially non-optimal stimuli. Response normalization after presentation of high contrasts does not affect burst length and is not affected by GABA.

### 1. THE REPRESENTATION OF VISUAL INFORMATION IN THE STRIATE CORTEX.

Most models of visual function are based on the notion that the visual world is internally represented by some form of linear deconstruction of the scene, into either specific features (e.g., Hubel & Wiesel, 1962; Bishop, Coombs & Henry, 1971) or spectral components (e.g., DeValois, DeValois & Yund, 1979). Both concepts have some attractive aspects, but resolution of the "correctness" of the competing schemes remains largely a theological issue. Common to both proposals is the notion that however the cells are parsing the scene, the characteristics of each detector (or filter) are both stationary and independent of one another. The latter requirement is implicit whenever the term "linear" is invoked, in that superposition, a fundamental requirement of linearity, cannot result if there is interaction between units.

Recent results from our laboratory suggest that in two fundamental domains, space and contrast, cortical cells interact, and are thus non-linear (Bonds, 1992). This means that the detector or filter characteristic of a cell can vary as the structure of the visual environment changes. A detector or filter characteristic that is measured in a specific way is thus strictly valid only under identical conditions of stimulation. The following presentation will briefly summarize evidence for the two classes of non-linearity mentioned above and will show that the interactions found in the two domains are apparently mediated by different mechanisms. A physiological mechanism that supports modulation of spike burst length has been found in the case of spatial interactions, but not in the case of contrast-dependent interactions, suggesting the existence of two specific pathways supporting non-linear cortical interdependencies.

### 2. METHODS.

In brief, single unit responses were recorded from Area 17 of cats that were paralyzed and anaesthetised with nitrous oxide and a barbiturate. Standard methods for preparation and recording were used (e.g., Bonds, 1989). Receptive fields were stimulated by drifting sinewave gratings generated on a CRT. Microiontophoresis of GABA and GABA blocking agents (*n*-methyl-bicuculline) was by means of a multi-barrelled pipette, with a sodium chloride channel for current compensation.

### 3. ORIENTATION-DEPENDENT INTERACTIONS.

Orientation selectivity is one of the most dramatic demonstrations of the filtering ability of cortical cells. While cells in the LGN are only mildly biased for stimulus orientation, cells in cortex are completely unresponsive to orthogonal stimuli and have tuning bandwidths that average only about 40-50° (e.g., Rose & Blakemore, 1974). How this happens remains

controversial, but there is general consensus that inhibition helps to refine orientation selectivity, although the schemes vary. The concept of *cross-orientation inhibition* (e.g., Morrone, Burr & Maffei, 1982) proposes that the inhibition is itself orientation selective and tuned in a complimentary way to the excitatory tuning of the cell, being weakest at the optimal orientation and strongest at the orthogonal orientation. More recent results, including those from our own lab, suggests that this is not exactly the case. We studied the orientation dependence of inhibition by presenting two superimposed gratings, a *base* grating at the optimal orientation to provide a steady level of background response activity, and a *mask* grating of varying orientation which yielded either excitation or inhibition that could supplement or suppress the base-generated response. There is some confusion when both base and mask generate excitation. In order to separate the response components from each of these stimuli, the two gratings were drifted at differing temporal frequencies, usually 2 vs 3 Hz. At least in simple cells, the individual contributions to the response from each grating could then be resolved by performing Fourier analysis on the response histograms.

Experiments were done on 52 cells, of which about 2/3 showed organized suppression from the mask grating (Bonds, 1989). Fig. 1B shows that while the mask-generated response suppression is somewhat orientation selective, it is by and large much flatter than would be required to account for the tuning of the cell. There is thus some orientation dependence of inhibition, but not specifically at the orthogonal orientation as might be expected. Instead, the predominant component of the suppression is constant with mask orientation, or *global*. This suggests that virtually any stimulus can result in inhibition, whether or not the recorded cell actually "sees" it. If any orientation-dependent component of inhibition is found, it is expressed in suppressive side-bands 20-30° from the peak orientation (filled circles, Fig 1C), which are likely to enhance any pre-existing orientation bias. Evidence for inhibition from nearby orientations is also found in cross-correlation studies (Hata, Tsumoto, Sato & Tamura, 1991).

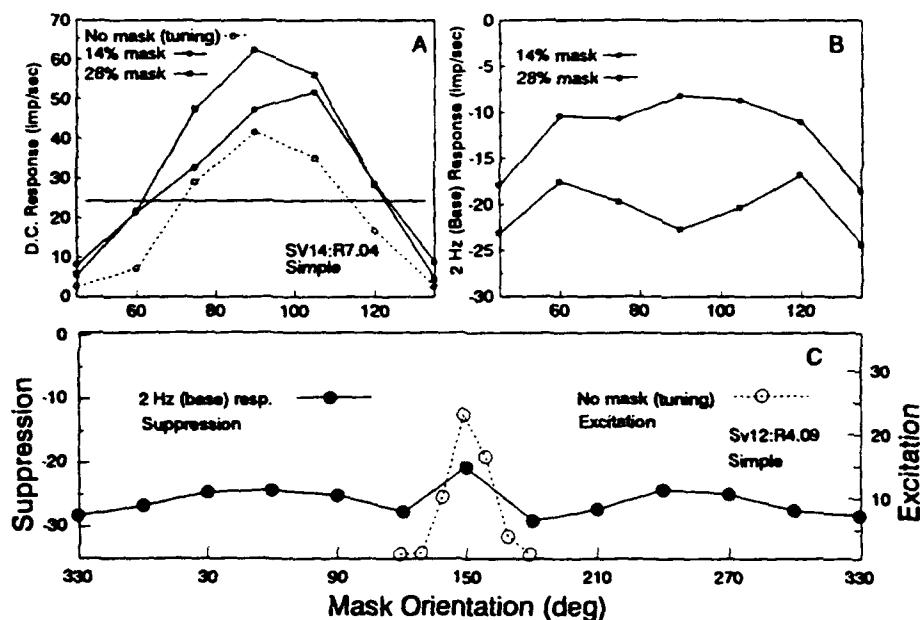


Fig. 1. Response suppression by mask gratings of varying orientation. (A) Impact of masks of 2 different contrasts on DC response component. The horizontal line denotes the response level from the base stimulus alone; the response is suppressed below this level at orientation extremes. The dotted line (open circles) is the orientation tuning curve of the cell. (B) Suppression of the 2 Hz (base-generated) response component, expressed by decrease (negative imp/sec) from response level arising from base stimulus alone. (C) Similar example for mask orientations spanning a full 360°.

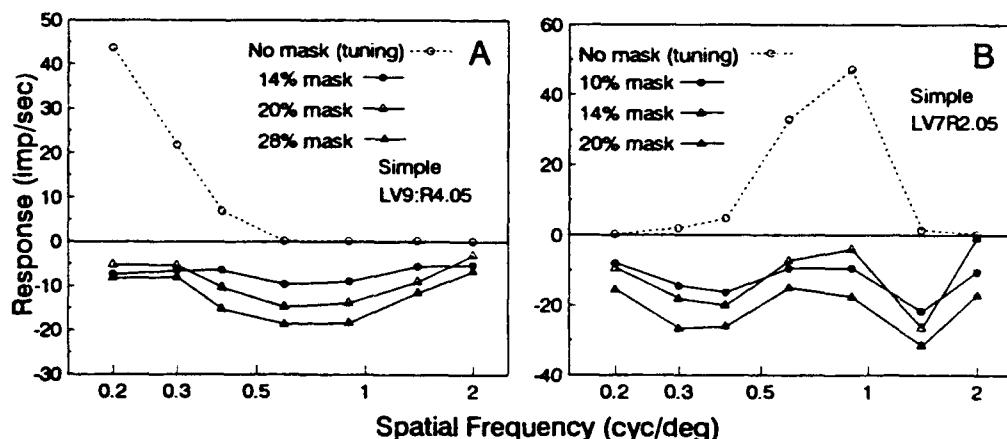
Thus the concept of cross-orientation inhibition is not particularly correct, since inhibition is found not just at the "cross" orientation but rather at all orientations, as also found by DeAngelis, Robson, Ohzawa & Freeman (1992). Even without orientation-selective inhibition, a scheme for establishment of true orientation selectivity from orientation-biased LGN input can be derived by assuming that the nonselective inhibition is graded and contrast-dependent and that it acts as a

thresholding device (Bonds, 1989).

The orientation selectivity of the inhibition is clearly not the same as that of the recorded cell, so it must be externally generated by other cells. Because this external signal can influence orientation selectivity of a given cell, the filter characteristic of that cell will depend on the stimulation of the pool that generates this external signal. This may not impact seriously on orientation selectivity when, e.g., comparing the results of grating vs bar stimulation, since both gratings and bars of similar orientation are likely to have about the same influence on both the recorded cell as well as its inhibitory pool. However, this gives no assurance that the orientation filtering of a cell measured with a grating is the same as the orientation filtering performed by the same cell when it is stimulated by a complex "natural" scene.

#### 4. SPATIAL FREQUENCY-DEPENDENT INTERACTIONS.

While most retinal and LGN cells are broadly tuned and predominantly low-pass, cortical cells generally have spatial frequency bandpasses of about 1.5-2 octaves (e.g., Maffei & Fiorentini, 1973). We have examined the influence of inhibition on spatial frequency selectivity using the same strategy as described above for orientation selectivity (Bauman & Bonds, 1991). A base grating, at the optimal orientation and spatial frequency, drove the cell, and a superimposed mask grating, at the optimal orientation but at different spatial and temporal frequencies, provided response facilitation or suppression. We defined three broad categories of spatial frequency tuning functions: Low pass, with no discernible low-frequency fall-off, band-pass, with a peak between 0.4 and 0.9 c/deg, and high pass, with a peak above 1 c/deg. About 75% of the cells showed response suppression organized with respect to the spatial frequency of mask gratings. For example, Fig. 2A shows a low-pass cell with high-frequency suppression and Fig. 2B shows a band-pass cell with mixed suppression, flanking the tuning curve at both low and high frequencies. In each case response suppression was graded with mask contrast and some suppression was found even at the optimal spatial frequency. This suggests again a component of global, or spatially non-specific, inhibition as was found in the orientation experiments described above. Some cells showed no suppression, indicating that the suppression was not merely a stimulus artifact. In all but 2 of 42 cases, the suppression was appropriate to the enhancement of the tuning function (e.g., low-pass cells had high-frequency response suppression), suggesting that the design of the system is more than coincidental. We believe that this phenomenon arises in the visual cortex, since no similar pattern of spatial frequency-dependent suppression was found in LGN cells.



**Fig 2.** Examples of spatial frequency-dependent response suppression. Upper broken lines show excitatory tuning functions and solid lines below zero indicate response reduction at three different contrasts. (A) Low-pass cell with high frequency inhibition. (B) Band-pass cell with mixed (low and high frequency) inhibition. Note suppression at optimal spatial frequency in both cases.

Again this result suggests a form of spatially-selective inhibition that serves to enhance filtering, here in the domain of spatial frequency. As in the case of orientation, the inhibition has spatial characteristics that differ clearly from those of the target cell, so it must arise in the network. This implies that spatial frequency selectivity of individual cells is likewise not a fixed property and will depend on the state of the network.

## 5. NON-STATIONARITY OF CONTRAST TRANSFER.

The two experiments described above demonstrate the existence of intrinsic cortical mechanisms that refine the spatial filter properties of the cells. They also reveal a *global* form of inhibition that is spatially non-specific. Since global inhibition is found even with spatially optimal stimuli, it can influence the form of the cortical contrast-response function, which is usually measured with optimal stimuli. This function is essentially logarithmic, with saturation or even super-saturation (a response downturn) at higher contrasts (e.g., Albrecht & Hamilton, 1982), as opposed to the more linear response behavior seen in cells earlier in the visual pathway. Cortical cells also show some degree of contrast adaptation; when exposed to high mean contrasts for long periods of time, the response vs contrast curves move rightward (e.g., Ohzawa, Sclar & Freeman, 1985). We addressed the question of whether contrast-response nonlinearity and adaptation might be causally related. In order to compensate for "intrinsic response variability" in visual cortical cells, experimental stimulation has historically involved presentation of randomized sequences of pattern parameters, the so-called multiple histogram technique (Henry, Bishop, Tupper & Dreher, 1973). Scrambling presentation order distributes time-dependent response variability across all stimulus conditions, but this procedure can be self-defeating by masking any influence that stimulus history might have on the response. We therefore presented cortical cells with ordered sequences of contrasts, first ascending then descending in a stepwise manner (Bonds, 1991). This revealed a clear and powerful response hysteresis. Fig. 3A shows a solid line representing the contrast-response function measured in the usual way, with randomized parameter presentation, overlaid on an envelope outlining responses to sequentially increasing or decreasing 3-sec contrast epochs; one sequential presentation set required 54 secs. Across 36 cells measured in this same way, the average response hysteresis measured at half maximum response amplitude corresponded to 0.36 log units of contrast. Some hysteresis was found in every cortical cell and in no LGN cells, so this phenomenon is intrinsically cortical.

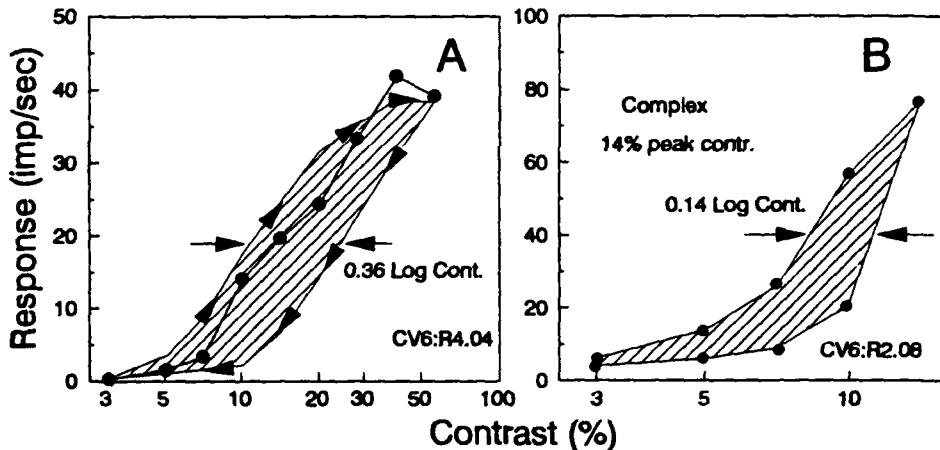


Fig. 3. Dynamic response hysteresis. A. A response function measured in the usual way, with randomized stimulus sequences (filled circles) is overlaid on the function resulting from stimulation with sequential ascending (upper level) and descending (lower level) contrasts. Each contrast was presented for 3 seconds. B. Hysteresis resulting from peak contrast of 14%; 3 secs per datum.

Hysteresis demonstrates a clear dependence of response amplitude on the history of stimulation: at a given contrast, the amplitude is always less if a higher contrast was shown first. This is one manifestation of cortical contrast adaptation, which is well-known. However, adaptation is usually measured after long periods of stimulation with high contrasts, and may not be relevant to normal behavioral vision. Fig. 3B shows hysteresis at a modest response level and low peak contrast (14%), suggesting that it is sufficiently sensitive to serve a major function in day-to-day visual processing, as well as influence the non-linearities found in the response vs contrast function. The speed of hysteresis also important to this issue, but it is not so easily measured. Some response histogram waveforms show consistent amplitude loss over a few seconds of stimulation (see also Albrecht, Farrar & Hamilton, 1984), but other histograms can be flat or even show a slight rise over time despite clear contrast adaptation (Bonds, 1991). This suggests the possibility that, in the classical pattern of any well-designed automatic gain control, gain reduction takes place quite rapidly, but its effects linger for some time.

The speed of reaction of gain change is illustrated in the experiment of Fig. 4. A "pedestal" grating of 14% contrast is introduced. After 500 msec, a contrast increment of 14% is added to the pedestal for a variable length of time. The response during the first and last 500 msec of the pedestal presentation is counted and the ratio is taken. In the absence of the increment, this ratio is about 0.8, reflecting the adaptive nature of the pedestal itself. For an increment of even 50 msec duration, this ratio is reduced, and it is reduced monotonically--by up to half the control level--for increments lasting less than a second. The gain control mechanism is thus both sensitive and rapid.

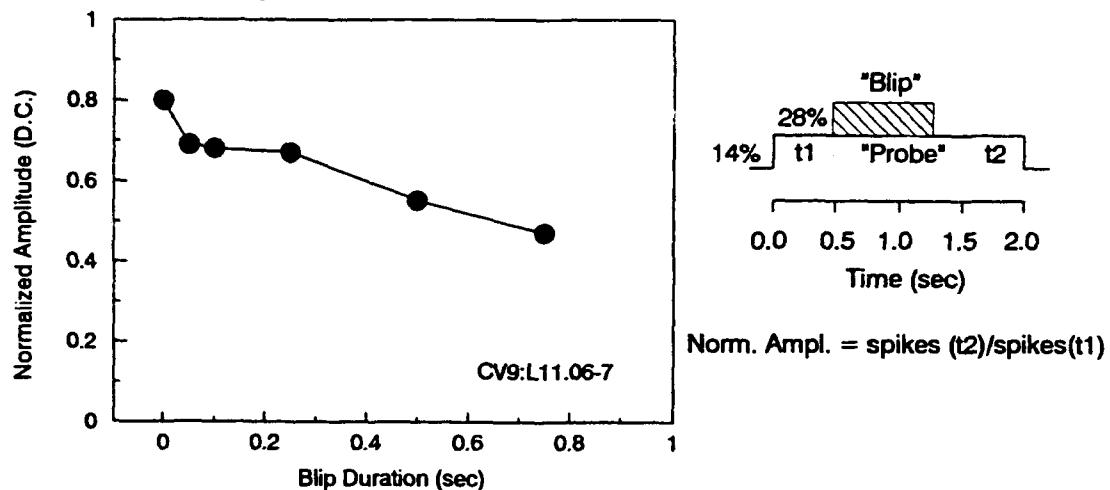


Fig 4. Speed of gain reduction. The ratio of spikes generated during the last and first 500 msec of a 2 sec pedestal presentation can be modified by a brief contrast increment (see text).

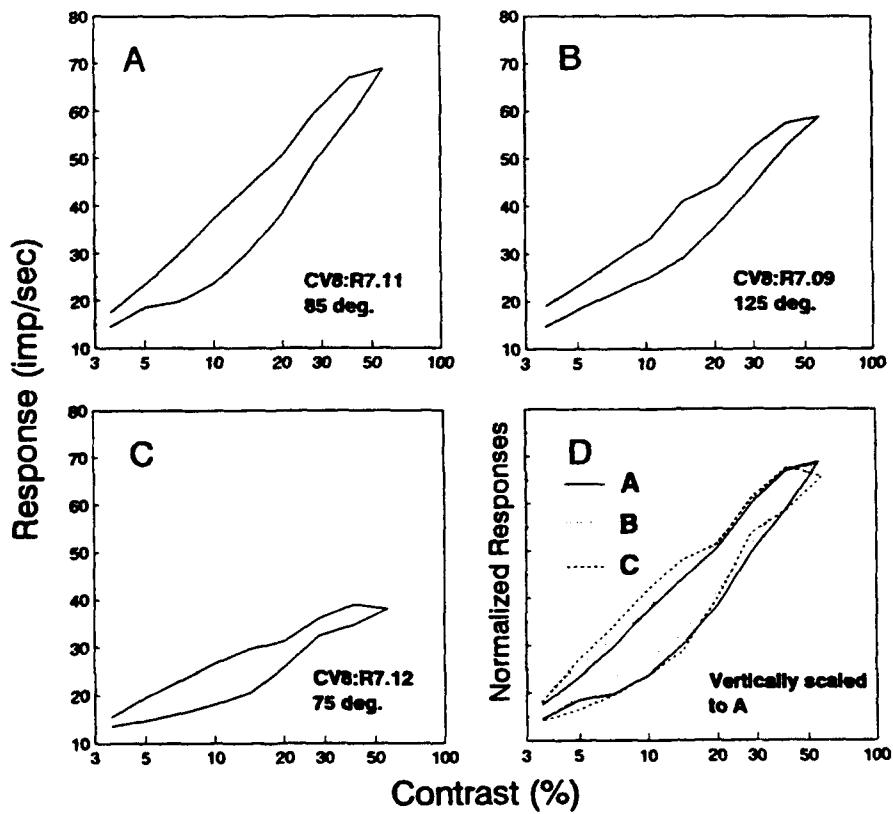


Fig 5. (A, B, C) Contrast hysteresis measured at three different orientations, to yield three different peak responses. Contrast is varied over the same range in each case. (D) When (B) and (C) are scaled to (A), the curves overlap, indicating little dependence of hysteresis on firing rate.

It is obvious that the time-dependent component of the contrast-response function results in a response non-linearity. What is less obvious is the "seriousness" of this non-linearity, at least with respect to its impact on the cortical coding of the visual scene. If the hysteresis were simply a consequence of the activity of the recorded cell, it would be expressed as a straightforward response compression and would not affect the integrative or spatial organization of the cell. If, however, the response hysteresis were some function of the activity of other cells as well, then the model of cortical cell function would become much more complicated due to the interactivity. This question was addressed in an indirect way by assessing the dependence of hysteresis on the activity of the recorded cell. Response hysteresis was measured with different orientations of the test stimulus, each yielding a different peak response level even though the set of contrasts used was identical in each case. Fig 5 shows that even though the peak response varies by more than 74% (from 39.9 to 69.5 imp/sec), when responses are normalized the degree of hysteresis is essentially identical, from 0.25 to 0.28 log units. This means that hysteresis is independent of the firing rate of the recorded cell and is rather a function of stimulus contrast. Like the spatial inhibition described above, contrast adaptation is not a property of the isolated cell and must depend on network activity.

Contrast adaptation can be thought of as a form of automatic gain control, since by moving the response vs log contrast curves to the right with higher ambient contrast it permits maintenance of a steep slope of the curve despite the limited dynamic range of the cell. The fact that it is controlled by network activity rather than the firing of the individual cell has an important implication. Its purpose is not to optimize contrast discriminability of the individual cell by keeping its firing constrained within a narrow range. If that were the case, then the contrast of the visual image, reflected by the contrast in firing between cells, would be minimized. Instead, it appears to be interactive so as to constrain the total activity of a regional group of cells within some limit. This provides not only contrast enhancement (through lateral inhibition) but also helps to prevent overdriving of cells at the next stage of visual processing, which could easily occur due to high signal convergence.

## 6. PHYSIOLOGICAL INDEPENDENCE OF TWO INHIBITORY MECHANISMS.

The experimental observations presented above demonstrate two basic phenomena: spatially-dependent and spatially-independent inhibition. Both are inherently non-linear processes, since they are capable of modifying the output of a given cell but are derived from signals generated by other cells. These processes challenge the notion that cortical representation of visual information is the result of static, linear filtering. One question that remains addresses the complexity of these cellular interactions: Are these two types of inhibition fundamentally different, or do they stem from the same physiological mechanism? This can be approached by examining the structure of a serial spike train generated by a cortical cell. In general, rather than being distributed continuously, cortical spikes are grouped into discrete packets, or bursts, with some intervening isolated spikes. Nearly all cortical cells are found to burst, and about half the cells we have recorded have 40% or more of their spikes contained in bursts. Burst behavior can be fundamentally characterized by two parameters: the burst frequency (bursts per second, or BPS) and the burst duration (measured in spikes per burst, or SPB). While we have analyzed cortical spike trains for these properties by using a number of approaches to define burst groupings, the simplest effective criterion is to consider spike intervals of 8 msec or less as belonging to bursts (see also Mandl, 1993). Other criteria change the following results quantitatively but not qualitatively.

Both burst frequency (BPS) and structure (SPB) depend strongly on mean firing rate, but once firing rate is taken into account, two basic patterns emerge. Consider two experiments on the same cell, both yielding firing rate variation about a similar range. In one experiment, firing rate is varied by varying stimulus contrast, while in the other, firing rate is varied by varying stimulus orientation. In all cases, burst frequency (BPS) is found to depend only on spike rate. In Fig. 6A, no systematic difference in BPS is seen between the experiments in which contrast (filled circles) and orientation (open squares) are varied. To quantify the difference between the curves, polynomials were fit to each and the quantity gamma, defined by the normalized (shaded) area bounded by the two polynomials, was calculated; here, it equalled about 0.03.

Fig. 6B shows that at similar firing rates, burst length (SPB) is markedly shorter when firing rate is controlled by varying orientation (open squares) as opposed to contrast (filled circles). In this pair of curves, the gamma (of about 0.25) is nearly ten times that found in Fig 6A. This is a clear violation of univariance, since at a given spike rate (output level), the structure of the spike train differs depending on the spatial configuration of the stimulus. Analysis of cortical response merely on the basis of overall firing rate thus does not give the signalling mechanisms the respect they are

properly due. This result also implies that the strength of signalling between nerve cells can dynamically vary independent of firing rate. Because of post-synaptic temporal integration, bursts of spikes with short interspike intervals will be much more effective in generating depolarization than spikes at longer intervals. Thus, at a given average firing rate, a cell that generates longer bursts will have more influence on a target cell than a cell that distributes its spikes in shorter bursts, all other factors being equal.

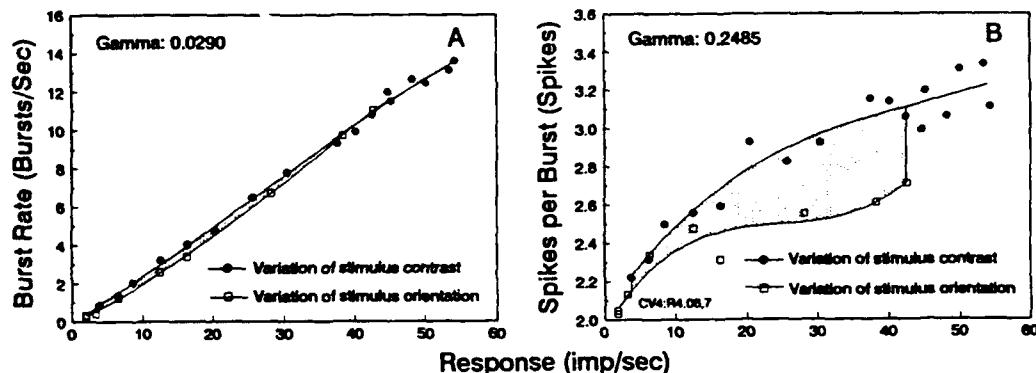


Fig 6. A. Comparison of burst frequency (bursts per second) as function of firing rate resulting from presentations of varying contrast (filled circles) and varying orientation (open squares). B. Comparison of burst length (spikes per burst) under similar conditions. Note that at a given firing rate, burst length is always shorter for non-optimal orientations. Shaded area (gamma) is quantitative indicator of difference between two curves.

This phenomenon was consistent across a population of 59 cells. Gamma, which reflects the degree of difference between curves measured by variation of contrast and by variation of orientation, averaged zero for curves based on number of bursts (BPS). For both simple and complex cells, gamma for burst duration (SPB) averaged 0.15.

At face value, these results simply mean that when lower spike rates are achieved by use of non-optimal orientations, bursts are shorter than when lower spike rates result from reduction of contrast (with the spatial configuration remaining optimal). While orientation manipulations result in inhibition that acts to change burst length, contrast manipulations appear to maintain a fixed relationship between firing rate and burst length. These results thus support the notion that there are at least two distinct forms of cortical inhibition, with unique physiological bases differentiated by the burst organization. The hypothesis that the two forms of inhibition are independent can be further tested by examining the behavior of spike bursts in the presence of explicit contrast-dependent inhibition. The model would predict that in experiments involving response hysteresis, in which only contrast is changed, burst duration would depend only on absolute firing rate.

Burst analysis was applied to seven cells tested for response hysteresis; representative results are summarized in Fig. 7. When hysteresis measures involve high contrasts and response saturation (Fig 7A), BPS remains a constant function of firing rate, but the SPB trajectory (filled circles, Fig 7B) is mixed. SPB rises with spike rate on the presentation of ascending contrast, but as the firing rate saturates, then super-saturates (falls with rising contrast) it drops dramatically and remains depressed over the remainder of the contrast presentations. This would initially suggest that the same mechanism that shortens bursts at non-optimal orientations also influences contrast gain control, which is at odds with the model of two independent inhibitory processes proposed above. However, a second manipulation on the same cell helps to clarify results. In this experiment (Fig 7C) the peak contrast was limited to 14%, which avoided any evidence of response saturation but still yielded significant response hysteresis. The SPB trajectory is exactly identical for both ascending and descending contrasts (Fig 7D), indicating that in this case the burst-shortening mechanism was not activated.

We can thus conclude that in the limited case of gain control invoked under conditions of moderate contrast not involving saturation, gain can be reduced without evidence of the inhibitory mechanism that shortens bursts. Equally interesting is the evidence showing that response saturation is a non-linearity that results from burst shortening, which results from the same inhibitory mechanism that helps to define spatial filter properties. (Note in Fig 7B that in the region of

supersaturation (contrasts above 14%), there is no significant change in the trajectory of the BPS (number of bursts) curve.) This suggests that saturation arises via inhibitory linkages from cells that have spatial tuning that is different from the recorded cell. These cells could then act as "lateral inhibitors" in the appropriate domain (orientation, spatial frequency) to sharpen the spatial selectivity of the recorded cell. Presentation of low-contrast stimuli optimal for the recorded cell (and thus non-optimal for the inhibitory cells) would not strongly activate the inhibitors, but higher contrasts would result in their recruitment, which could then result in burst shortening that yielded saturation.

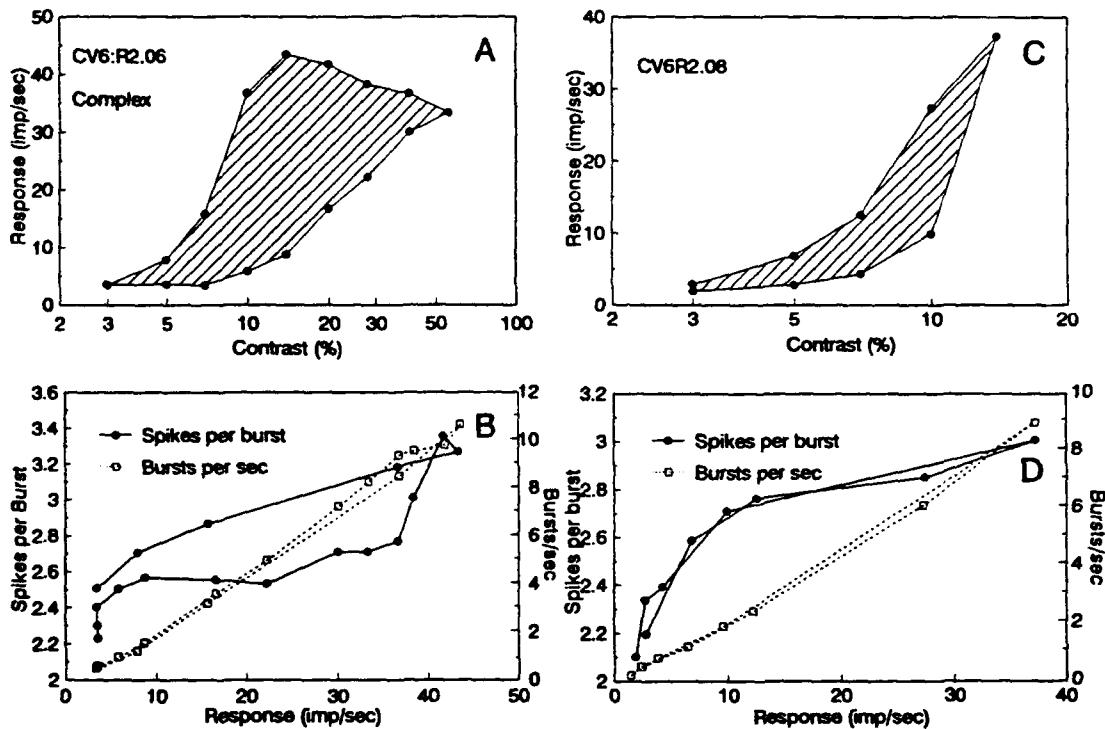


Fig 7. (A, C) Contrast hysteresis curves measured on a complex cell; 3 seconds per datum. In A the peak contrast is 56%, whereas in C it is 14%. (B,D) BPS and SPB results from (A) and (C), respectively. In (B), SPB falls dramatically as the response vs contrast function drops at higher contrasts, and remains depressed. In (D), the SPB trajectory is identical for rising and falling contrasts (and firing rates), even though substantial contrast adaptation is seen.

## 7. NEUROTRANSMITTER BASIS FOR BURST MODULATION.

Microiontophoretic injection of bicuculline, a GABA blocker, has been found to broaden orientation selectivity in some cells (e.g., Sillito, 1975, 1979). The active suppression of cortical responses resulting from presentation of stimuli at non-optimal orientations is thus likely to be mediated by GABA. A causal link is possible between activation of GABA-mediated suppressive mechanisms and the shortening of bursts, since both occur when stimuli of non-optimal orientations are presented. This linkage is also suggested by the observation of burst shortening in the presence of GABA and burst lengthening in the presence of bicuculline found in cat somatosensory cortex (Dykes, Metherate, Landry & Hicks, 1984).

To test this hypothesis, we performed burst analysis on spike trains under control conditions and in the presence of microiontophoretically injected GABA and bicuculline. We confirmed that burst length (controlled for firing rate) is reduced by GABA and increased by (n-Methyl-)bicuculline. Experiments in which GABA was injected were problematic. We wanted a graded reduction of responsiveness, but even with a dilute solution many cells stopped firing in the presence of GABA. For this reason, control and GABA-influenced response measurements were completed in only nine cells. In each case, burst length at a given firing rate was reduced (Fig 8A). To provide a quantitative basis for the effect, the burst length vs firing rate curves were fit with simple linear regression. The burst length was then estimated at half the maximum firing rate of the cell (vertical line, Fig 8A) for both the control and GABA conditions. A histogram of the

results is seen in Fig 8C. The mean of the distribution is reduced 13.8% in the presence of GABA.

Injection of n-Methyl-bicuculline results in higher firing rates, so in this case we had a sample of 66 trials (29 cells). In all but two trials, burst length increased at constant firing rate from the control condition (Fig 8D). The histogram of burst lengths at 50% maximum firing rate, constructed in the same way as described above, shows a mean increase of the distribution of 39% in the presence of n-Methyl-bicuculline. These observations are consistent with the idea that the shortening of bursts found when cells are stimulated by non-optimal orientations is an indicator of GABA-mediated inhibition.

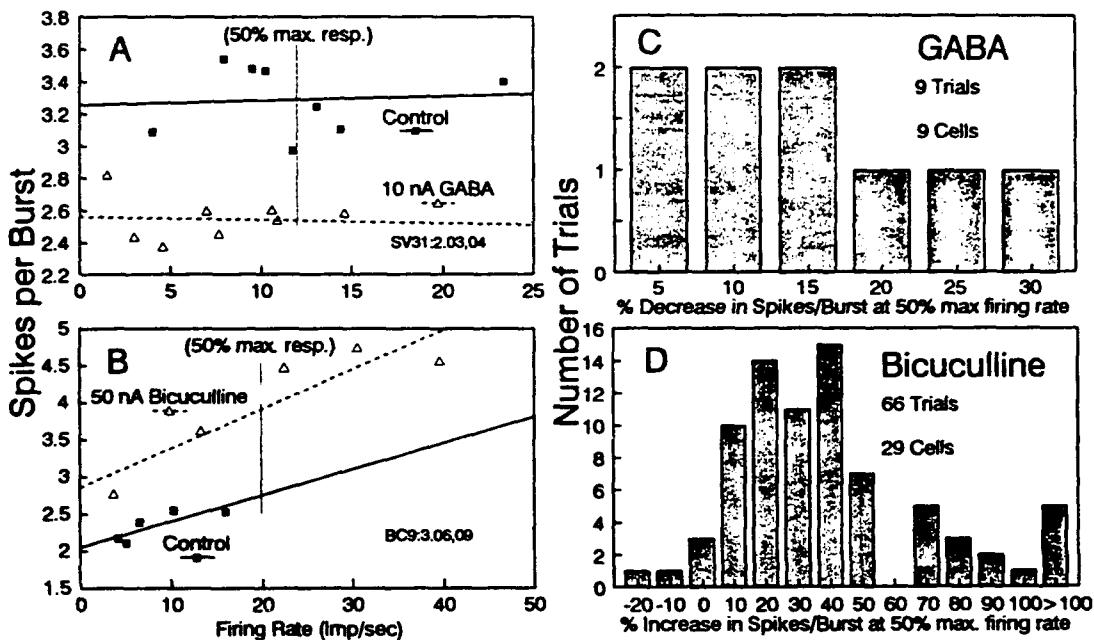


Fig 8. (A, B) Impact of GABA and bicuculline (respectively) on SPB of cortical discharges. Dark squares (with solid regression line) represent control data, open triangles (with dotted regression line) represent pharmacological influence. (C) Distribution of SPB decrease with GABA. (D) Distribution of SPB increase with bicuculline.

## 8. DISCUSSION.

Taken together, these results support the existence of at least two distinct forms of cortical inhibition, with unique physiological bases differentiated by the burst organization in cortical spike trains. One type of inhibition is (1) spatially selective, (2) results in burst shortening when activated and (3) is influenced by GABA. The other type (1) does not appear to be spatially selective, (2) does not exhibit burst shortening when activated and (3) is not influenced by GABA (DeBruyn & Bonds, 1986). The first form of inhibition can be thought of as providing lateral inhibition in the domains of orientation and spatial frequency, so as to refine the spatial selectivity of the cell. At least in the orientation domain, the visual cortex is organized in columns, with adjacent columns having nearly identical orientations (e.g., Hubel & Wiesel, 1962). Since the strongest "orientation-selective" inhibition appears to come from nearby orientations (Fig 1; Hata, et al., 1991), this would suggest circuitry that is local and, from the results of section 7, GABA-ergic. The architectural organization of spatial frequency is less well-understood, but one might conjecture that spatial frequency-dependent inhibition is mediated through the same pathways.

The substrate of the contrast gain control is less apparent. There is a component of inhibition that is found at all orientations and spatial frequencies visible to the cat (Bonds, 1989; DeAngelis, et al., 1992) that may be an expression of this mechanism. While contrast adaptation is found at many orientations, the result of Fig 5 does not necessarily support the notion that it is equally effective at all orientations. Experiments with a single grating cannot resolve this issue since no response is detectable outside the orientation tuning curve. What we do know is that contrast adaptation

can be found without burst shortening and that it is not influenced by GABA. Since it is not simply dependent on the firing rate of the individual cell, a second inhibitory pathway must exist that works in combination with that described above.

Both types of inhibition are capable of introducing substantial spatial and/or point non-linearities to the response of a given cortical cell. Under the usual experimental conditions, stimuli are composed of simple features or are spectrally pure, and are configured as "optimal" (across the limited range of parameters that is usually varied) for the cell under study. Activation of the contrast gain control can account for some of the contrast non-linearities seen under "optimal" stimulation, but as seen in Fig 7, at high contrasts involving saturation and super-saturation, the burst-shortening component of inhibition also appears. For non-optimal stimulation, e.g. at less effective orientations, burst-shortening is seen to have even more (pressive) impact on the response. This is presumably because an orientation that is non-optimal for one cell can be more optimal for its neighbor, and nearby cells can be mutually inhibitory through GABAergic pathways. The balance between inhibitory influences thus varies with the spatial organization of the stimulus. Under natural viewing conditions, one must expect a broad distribution of orientations and feature (or spatial frequency) varieties that, unlike most experiments, simultaneously activate many or all cells. Because of interactive influences between cells, cellular filter characteristics under these conditions will be dynamic, depending on the nature of the scene viewed, and will certainly differ from those found experimentally (Gilbert & Wiesel, 1990).

The extent of the influences of interdependencies on filter properties remains unknown, as well as the rationale for their existence. One might guess that spatially dependent inhibition serves to enhance filter selectivity and that contrast gain control eases overloading of subsequent stages and helps the signal-to-noise ratio, but these presumptions are founded on our understanding of the mechanisms as studied with simple stimuli. A more complete understanding will only come from more complex and realistic stimulation paradigms.

## 9. ACKNOWLEDGEMENTS

This work was done in collaboration with Ed DeBruyn, Lisa Bauman and Brian DeBusk. Supported by NIH RO1-EY03778 and NIH Core Grant R30-EY08126..

## 10. REFERENCES

D.G. Albrecht & D.B. Hamilton, "Striate cortex of monkey and cat: contrast response functions," *Journal of Neurophysiology* **48**, 217-237, 1982.

D.G. Albrecht, S.B. Farrar & D.B. Hamilton. "Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex," *Journal of Physiology* **347**, 713-739, 1984 .

L.A. Bauman & A.B. Bonds. "Inhibitory refinement of spatial frequency selectivity in single cells of the cat striate cortex," *Vision Research* **31**, 933-944, 1991.

P.O. Bishop, J.S. Coombs. & G.H. Henry. "Responses to visual contours: Spatio-temporal aspects of excitation in the receptive fields of simple striate neurons," *Journal of Physiology* **219**, 625-657, 1971.

A.B. Bonds. "The role of inhibition in the specification of orientation selectivity of cells of the cat striate cortex," *Visual Neuroscience* **2**, 41-55, 1989.

A.B. Bonds. "Temporal dynamics of contrast gain control in single cells of the cat striate cortex," *Visual Neuroscience* **6**, 239-255., 1991.

A. B. Bonds. "Spatial and temporal non-linearities in the receptive fields of striate cortical cells," in Non-Linear Vision, eds. Pinter & Nabet, pp. 329-352. CRC Press Cybernetics Series, Boca Raton, 1992.

E.J. DeBruyn & A. B. Bonds. "Contrast adaptation in cat visual cortex is not mediated by GABA," *Brain Research* **383**,

339-342.

G.C. DeAngelis, J.G. Robson, I. Ohzawa & R.D. Freeman. "Organization of suppression in receptive fields of neurons in cat visual cortex," *Journal of Neurophysiology* **68**, 144-163, 1992.

K.K. DeValois, R.L. DeValois & E.W. Yund. "Responses of striate cortex cells to gratings and checkerboard patterns." *Journal of Physiology* **291**, 483-505, 1979.

R.W. Dykes, P. Landry, R. Metherate & T.P. Hicks. "Functional role of GABA in cat primary somatosensory cortex: Shaping receptive fields of cortical neurons," *Journal of Neurophysiology* **52**, 1066-1093, 1984.

C.D. Gilbert & T.N. Wiesel. "The influence of contextual stimuli on the orientation selectivity of cells in the primary visual cortex of the cat," *Vision Research* **30**, 1689-1701, 1990.

Y. Hata, T.Tsumoto, H.Sato & H. Tamura. "Horizontal interactions between visual cortical neurones studied by cross-correlation analysis in the cat," *Journal of Physiology* **441**, 593-614, 1991.

G. Henry, P.O. Bishop, R.M. Tupper & B. Dreher. "Orientation specificity of cells in cat striate cortex," *Vision Research* **13**, 1771-1779, 1973.

D.H. Hubel & T.N. Wiesel. "Receptive fields, binocular interaction andfunctional architecture in the cat's visual cortex", *Journal of Physiology* **160**, 106-154. 1962.

L. Maffei & A. Fiorentini. "The visual cortex as a spatial frequency analyzer," *Vision Research* **13**, 1255-1267, 1973.

G. Mandl. "Coding for stimulus velocity by temporal patterning of spike discharges in visual cells of cat superior colliculus." *Vision Research* **33**, 1451-1476, 1993.

M.C. Morrone, D.C. Burr & L. Maffei. "Functional implications of cross-orientation inhibition of cortical visual cells," *Proceedings of the Royal Society (London) B* **216**, 335-354, 1982.

I. Ohzawa, G. Sclar & R.D. Freeman. "Contrast gain control in the cat's visual system," *Journal of Neurophysiology* **54**, 651-667, 1985.

D.Rose & C.B. Blakemore. "An analysis of orientation selectivity in the cat's visual cortex," *Experimental Brain Research* **20**, 1-17, 1974.

A.M. Sillito. "The effectiveness of bicuculline as an antagonist of GABA and visually evoked inhibition in the cat's striate cortex," *Journal of Physiology* **250**, 287-304, 1975.

A.M. Sillito. "Inhibitory mechanisms influencing complex cell orientation selectivity and their modification at high resting discharge levels," *Journal of Physiology* **289**, 33-53, 1979.

**Visual cortex neurons in monkey and cat:  
Contrast response nonlinearities and stimulus selectivity.**

Duane G. Albrecht  
and  
Wilson S. Geisler

Department of Psychology  
University of Texas  
Austin, Texas 78712

**ABSTRACT**

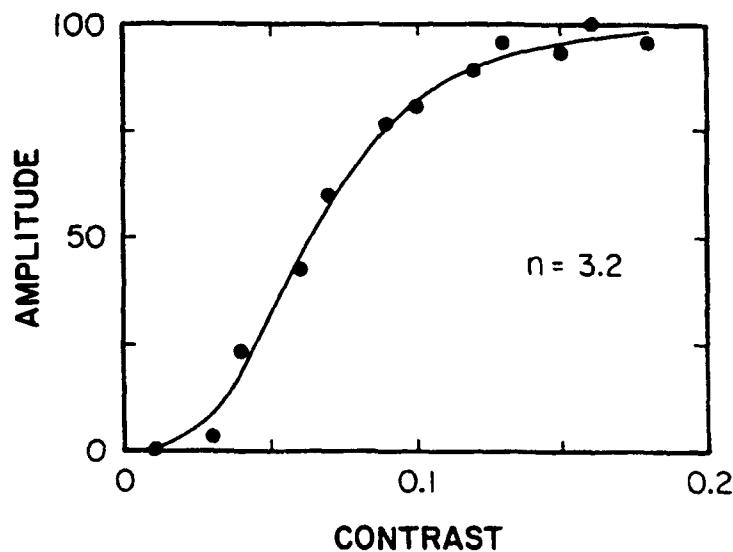
The contrast response functions of cat and monkey visual cortex neurons reveal two important nonlinearities: expansive response exponents and contrast gain control. These two nonlinearities (when combined with a linear spatiotemporal receptive field) can have beneficial consequences on stimulus selectivity. Expansive response exponents enhance stimulus selectivity introduced by previous neural interactions, thereby relaxing the structural requirements for establishing highly selective neurons. Contrast gain control maintains stimulus selectivity, over a wide range of contrasts, in spite of the limited dynamic response range and the steep slopes of the contrast response function.

**1. CONTRAST RESPONSE NONLINEARITIES**

Neurons in the visual cortex of monkeys and cats are known to be quite selective along a variety of stimulus dimensions: spatial position, spatial frequency, temporal frequency, orientation, direction of motion, etc.<sup>1-3</sup> Each neuron is analogous to a narrow-band filter, which transmits and signals only a limited portion of the available information. It seems reasonable to suppose that this type of stimulus selectivity plays a fundamental role in the overall process of vision and that one goal of the neural machinery is to produce individual neurons tuned to specific visual qualities. It is clear

that some of this selectivity is acquired through linear summation of inputs as described by the spatiotemporal receptive field, or the spatiotemporal transfer function.<sup>4-9</sup> However there are nonlinear interactions that also contribute to this goal. Specifically, the contrast response functions of neurons recorded from area 17 in the visual cortex reveal two nonlinearities, expansive response exponents and contrast gain control, that have rather beneficial consequences with respect to establishing and maintaining stimulus selectivity.

A typical contrast response function is illustrated in Figure 1. The responses of a representative neuron recorded from the striate visual cortex of a monkey are plotted as a function of the contrast of an optimal spatial frequency grating pattern. As the contrast of the grating increases from zero, the response increases rapidly with a power function exponent greater than 1.0; that is, when plotted on log coordinates the slope is greater than 1.0. The value of this expansive exponent varies from cell to cell; for this particular cell the exponent was 3.2; for some cells it exceeds 5.0; the average value is approximately 2.5. The rapidly accelerating response rate soon saturates at a maximum value. These facts have been established through measurements on many hundreds of neurons, in both cat and monkey, in several different laboratories.<sup>9-14</sup>



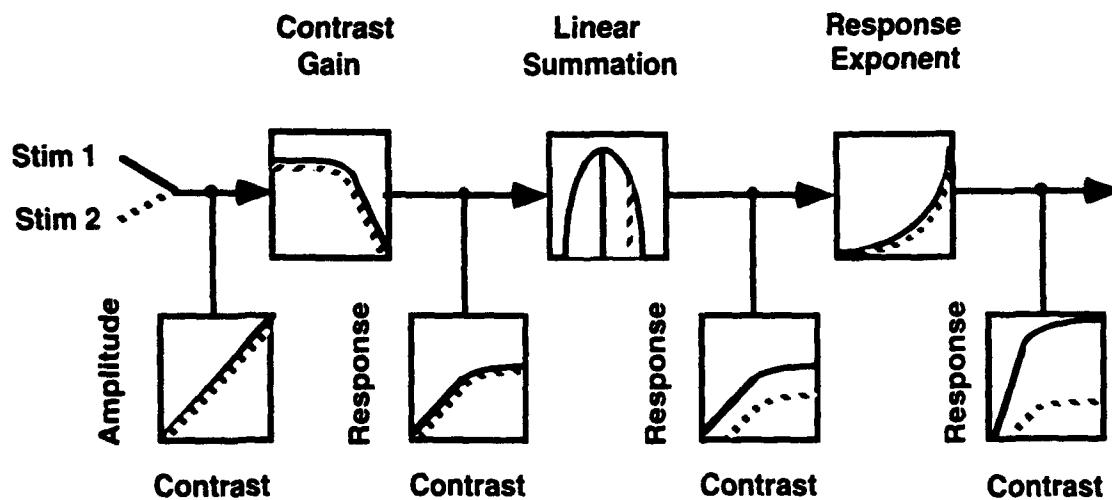
**Figure 1.** Contrast response function of a typical visual cortex neuron. The smooth curve through the data points is the best fit of a saturating power function. This cell illustrates the two nonlinearities: as the contrast increases, the response increases rapidly (in this case with an expansive power function exponent of 3.2) and then saturates.

The properties of the saturation nonlinearity, seen in the contrast response functions of cortical cells, turned out to be rather interesting and certainly unexpected. By measuring the contrast response function at different spatial frequencies, we found that the saturation is not really tied to the overall level of the response, per se; rather, the saturation is determined by the overall level of the contrast.<sup>10</sup> This fact, and other recent evidence<sup>13-16</sup> lead to the conclusion that the saturation is due to a fast-acting, multiplicative, contrast gain control mechanism.

## 2. CONTRAST-GAIN/EXPONENT MODEL

We have developed a formal model, the *contrast-gain/exponent (CGE) model* which incorporates these two nonlinearities into the established notion of linear spatiotemporal filtering.<sup>14</sup> This model is composed of three basic components: a linear spatiotemporal filter, contrast gain control, and an expansive response exponent. Heeger has developed and tested a similar model.<sup>17-20</sup> Figure 2 illustrates the consequences of these various operations for an optimal stimulus and a nonoptimal stimulus. On the input side, the amplitude increases linearly with contrast and is equal for the two stimuli. After the contrast gain is applied, the saturation nonlinearity becomes apparent, although the two equal contrast stimuli continue to evoke equal responses. After the linear filter, the responses are no longer equivalent; the responses to the nonoptimal stimulus are shifted down, by a constant ratio, at all contrasts. Finally, after the expansive response exponent is applied, the differences in the responses to the two stimuli are greatly magnified.

### Contrast-Gain Exponent Model

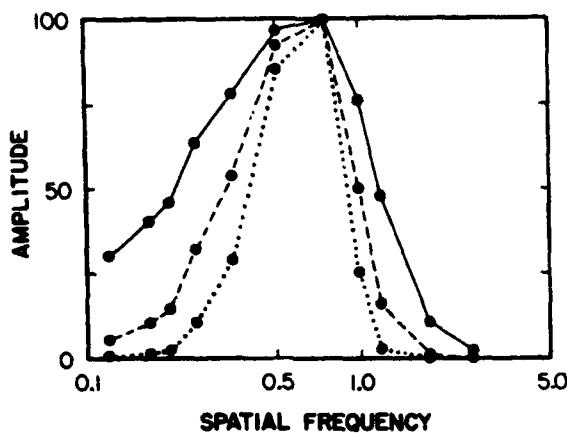


**Figure 2.** The effect of each component of the contrast-gain/exponent model on the responses to an optimal stimulus (solid line) and a nonoptimal stimulus (dashed line). The three mechanisms are illustrated symbolically in the upper row of boxes. The resulting contrast response functions before and after each stage are shown in the lower row of boxes.

The basic idea is that the linear filter establishes a certain degree of stimulus selectivity through simple linear summation of inputs; the shape of the spatiotemporal receptive field produces a certain degree of orientation selectivity (due to the length of the receptive field), spatial frequency selectivity (due to the width, strength, and number of parallel excitatory and inhibitory regions), direction selectivity (due to the orientation in space/time), etc. If the output of the linear filter is passed through an expansive nonlinearity, the stimulus selectivity along all these dimensions will be greatly enhanced. Contrast gain control will ensure that these selectivities are maintained across a broad range of contrasts, in spite of the steep slope and the limited dynamic range of the contrast response.

### 3. EXPONENT ENHANCES SELECTIVITY

Consider the effect of an expansive exponent on stimulus selectivity. The basic effect of the exponent is to enhance stronger excitations disproportionately, relative to weaker excitations; and thus, optimal stimuli are disproportionately enhanced relative to nonoptimal stimuli.

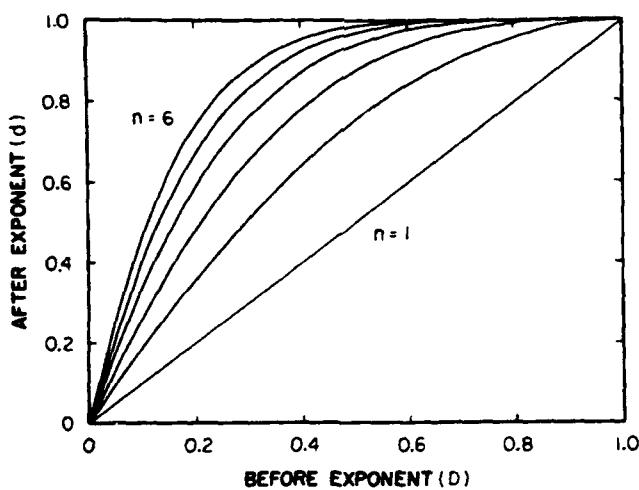


**Figure 3.** Effect of expansive exponent on spatial frequency selectivity. The filled circles connected by the solid lines plot the responses of a typical cat LGN cell (bandwidth=2.5 octaves). The filled circles connected by the dashed line plot the effect of applying an expansive exponent of 2.5 (bandwidth=1.5 octaves). The circles connected by the dotted line plot the effect of an exponent of 5.0 (bandwidth=1.0 octaves).<sup>14</sup>

Figure 3 illustrates the potential effect of an expansive exponent on spatial frequency selectivity. The circles connected by the solid line plot the responses of an LGN cell as a function of spatial frequency (the outermost curve). While this cell does show a certain degree of selectivity, due to its center/surround receptive field, it is rather broadly tuned --

it is not very selective for spatial frequency. It has a bandwidth of approximately 2.5 octaves. The circles connected by the dashed line illustrate what would happen to the responses if they were passed through a cell with an expansive exponent of 2.5; the circles connected by the dotted line illustrate the effect of an exponent of 5.0 (the innermost curve). As can be seen, the expansive exponent produces considerable narrowing or sharpening of the spatial frequency selectivity. An exponent of 2.5 reduced the bandwidth to approximately 1.5 octaves; an exponent of 5.0 reduced the bandwidth to approximately 1.0.

Figure 4 illustrates the effect of expansive exponents on the degree of direction selectivity. The direction selectivity of a strictly linear filter is plotted along the x-axis; the direction selectivity following expansion due to a power function exponent is plotted along the y-axis. The smooth curves show the effect of exponents ranging from one (the diagonal line -- no effect) through 6. The curves illustrate that expansive exponents can substantially increase the direction selectivity. For example, if the direction selectivity index were 0.3 before expansion, it is more than doubled by an exponent of 3.0. The exponent can make a very direction selective filter from a linear filter that is only partially direction selective.



**Figure 4.** Effect of expansive exponents on direction selectivity: each line plots the direction selectivity before (horizontal axis) and after (vertical axis) applying exponents ranging from 1.0 through 6.0. Direction selectivity is defined as  $(R_p - R_n) / R_p$ , where  $R_p$  is the magnitude of response in the preferred direction and  $R_n$  is the magnitude of response in the nonpreferred direction.<sup>14</sup>

#### 4. EXPONENTS EASE STRUCTURAL REQUIREMENTS

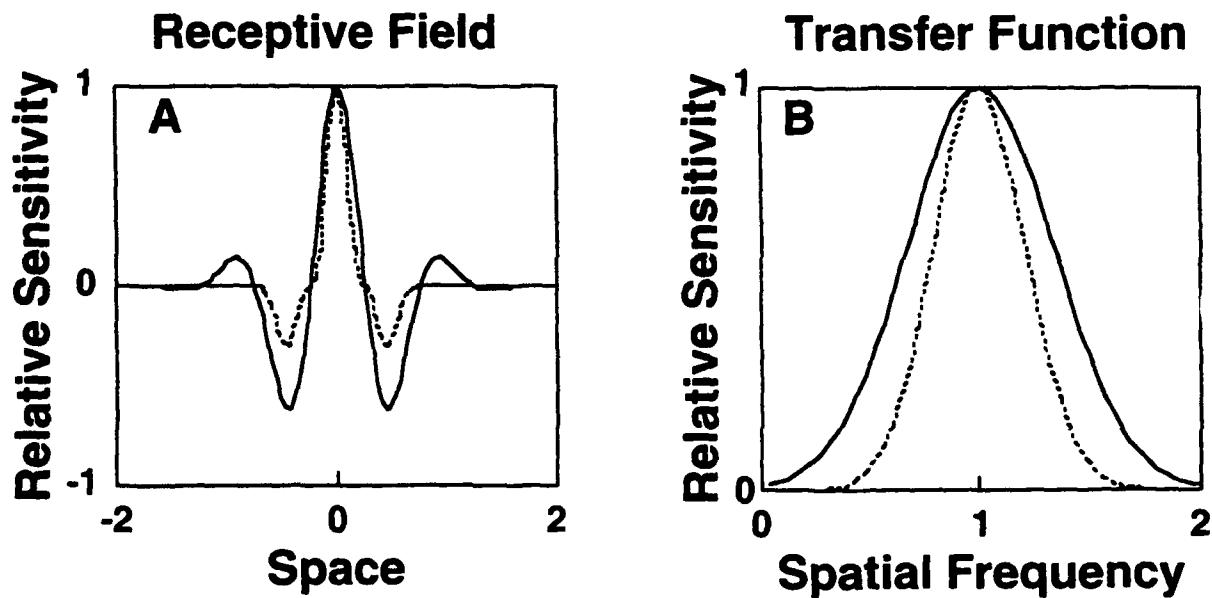
Expansive exponents can potentially ease the structural requirements for producing highly selective neurons. Consider spatial frequency selectivity. Given strict linearity, as the number of parallel excitatory and inhibitory regions within the receptive field is increased, the spatial frequency selectivity is increased. Thus, in order to produce a linear cell with a high degree of spatial frequency tuning, the spatial receptive field must be composed of many flanking excitatory and inhibitory regions. For a cell to have a bandwidth of say 0.8 octaves the receptive field would have approximately eight spatially antagonistic regions. On the other hand, if an expansive exponent is introduced after the linear filter, the number of antagonistic regions can be reduced. An expansive exponent of 2.5 could decrease the number by a factor of approximately two.

Consider orientation selectivity. Given strict linearity, as the length of the receptive field is increased, the orientation tuning is increased. Thus, in order to produce a linear cell with a high degree of orientation tuning, the spatial receptive field must be very long. On the other hand, if an expansive exponent is introduced after the linear filter, then the length can be reduced. Again, an exponent of 2.5 could decrease the required length by a factor of approximately two.

Consider direction selectivity. Given strict linearity, as the strength of the oriented component of the receptive field in the space-time domain is increased, the direction selectivity is increased. Thus, in order to produce a linear cell with a high degree of direction selectivity, the cell must be very strongly oriented in space time. On the other hand, if an expansive exponent is introduced after the linear filter, the strength of space-time orientation can be reduced. DeAngelis et al. have recently demonstrated that the discrepancy between the measured spatiotemporal RF and the measured degree of direction selectivity was diminished when the effects of the measured exponent were taken into consideration.<sup>9</sup>

The effect of the response exponent can potentially help account for a number of discrepancies between the degree of stimulus selectivity and the exact shape of the receptive field. In general, the degree of spatial frequency selectivity, orientation selectivity, and direction selectivity are greater than what would be expected from the shape of the receptive field. In the past, we and others have proposed various factors which could potentially account for the lack of correspondence and the increased stimulus selectivity; for example, various kinds of inhibitions from nonoptimal stimuli. In fact, the increased selectivity may well be a simple consequence of the expansive exponents seen in the contrast response functions.

Figure 5 attempts to illustrate this point. The solid line on the left side of this figure is the spatial receptive field profile (more accurately, the impulse response) of a given linear filter; the solid line on the right is the spatial frequency selectivity of the linear filter. These two constitute a transform pair -- the space and spatial frequency representations of a hypothetical strictly linear filter. They are Fourier transforms of each other. If the output of this strictly linear filter is passed through an expansive exponent of say 2.5 then both the measured receptive field and the measured spatial frequency tuning would change -- as illustrated by the superimposed dashed lines. The bandwidth of the spatial frequency tuning becomes narrower. The stronger input from the optimal frequencies is disproportionately enhanced relative to the weaker input from the nonoptimal frequencies. Similarly, the width of the receptive field becomes narrower. The stronger responses from the center would be disproportionately enhanced relative to the weaker responses from the other flanking regions.



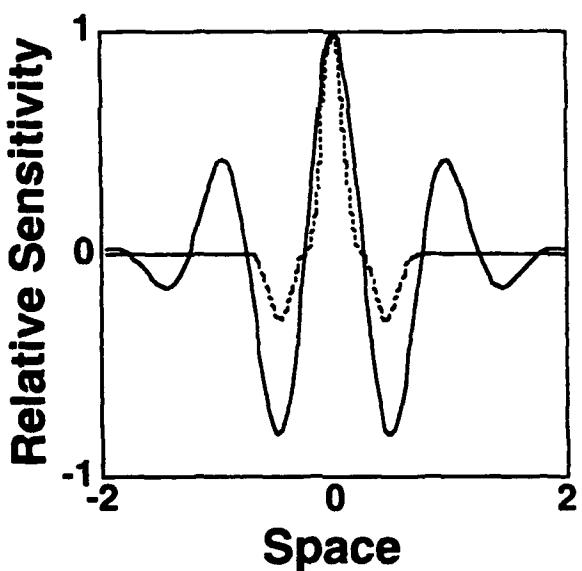
**Figure 5.** Spatial receptive field profile(A), and the corresponding spatial frequency tuning (B), before and after an expansive exponent. The solid lines plot the linear pair and the dashed lines plot the exponentiated pair.

These dashed lines would constitute the measured receptive field and the measured spatial frequency tuning of this filter, after exponentiation. Note that they are no longer a Fourier transform pair. The measured receptive field, following exponentiation, does not appear to have enough flanking regions to account for the narrow spatial frequency tuning. The exponent has simultaneously attenuated the weaker responses from nonoptimal/peripheral flanking regions and the weaker responses from nonoptimal/peripheral spatial frequencies. The expansive exponent has increased the

localization in both space and spatial frequency. The exponent has decreased the bandwidth of the spatial frequency tuning from 1.2 octaves to 0.7 octaves -- nearly a factor of two.

In order to produce spatial frequency tuning of 0.7 octaves using strictly linear mechanisms, the spatial receptive field profile would have to contain many flanking regions of excitation and inhibition -- as shown in Figure 6. The solid line is the receptive field profile of a linear mechanism with a spatial frequency tuning of 0.7 octaves. The superimposed dashed line is taken from Figure 5a -- it is the exponentiated receptive field that would correspond to the 0.7 octaves exponentiated spatial frequency tuning. Over the past few decades, many different laboratories have noted that the receptive fields of narrowly tuned cells generally do not have the receptive field expected from strictly linear mechanisms. The contrast response exponents can potentially help account for some of the these longstanding discrepancies.

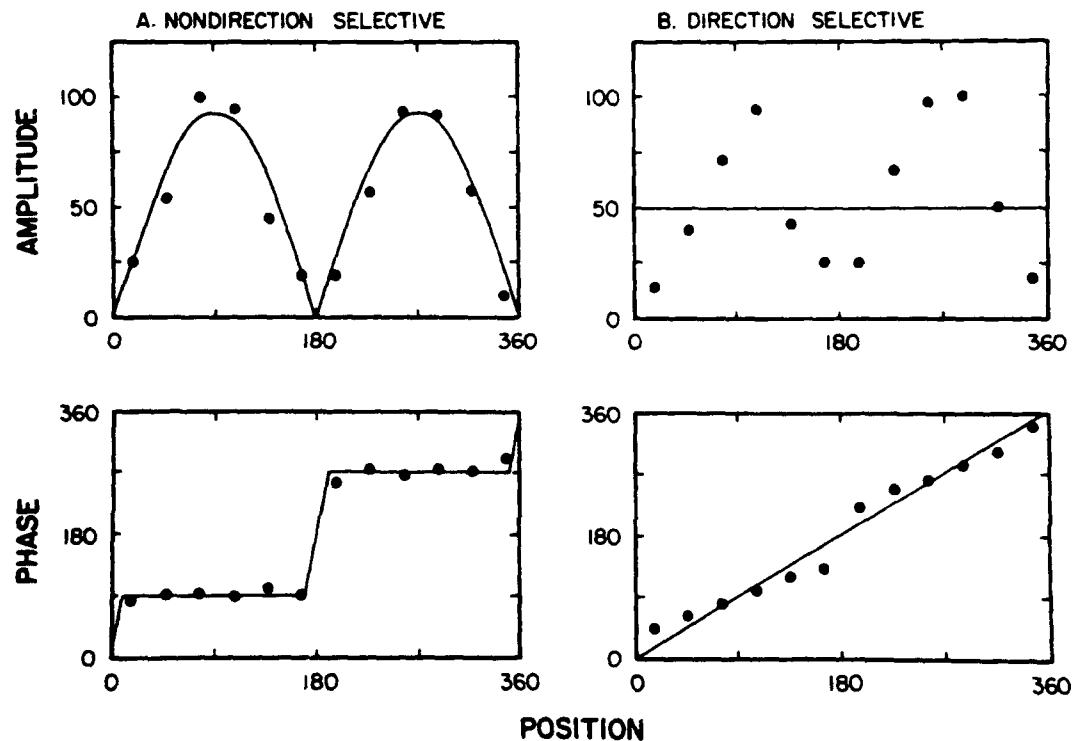
### Predicted Mismatch of Linear Analysis



**Figure 6.** Spatial receptive profile for a linear filter (solid line) and exponentiated filter (dashed line). The resulting spatial frequency tuning for both of these RFs is 0.7 octaves.

There is some evidence to support the above propositions. We have examples of individual cells which illustrate that the expansive exponent can help reconcile differences between the measured selectivity and the measured receptive field, and we are in the process of measuring and assessing the generality of the propositions for a large population of neurons. For example, we have completed one study of direction selectivity which clearly demonstrates the effects of the response exponent. We and

others<sup>21,22</sup> had previously noted that the measured direction selectivity of cortical cells was greater than what would be expected from the measured responses to stationary flashing stimuli, if only linear summation of inputs is taken into account. As Reid et al. stated: "only about half of the direction selectivity could be accounted for on the basis of linear mechanisms (p. 8742)." However, if the effects of the expansive exponent are taken into account, then the measured direction selectivity is consistent with the measured responses to stationary stimuli.

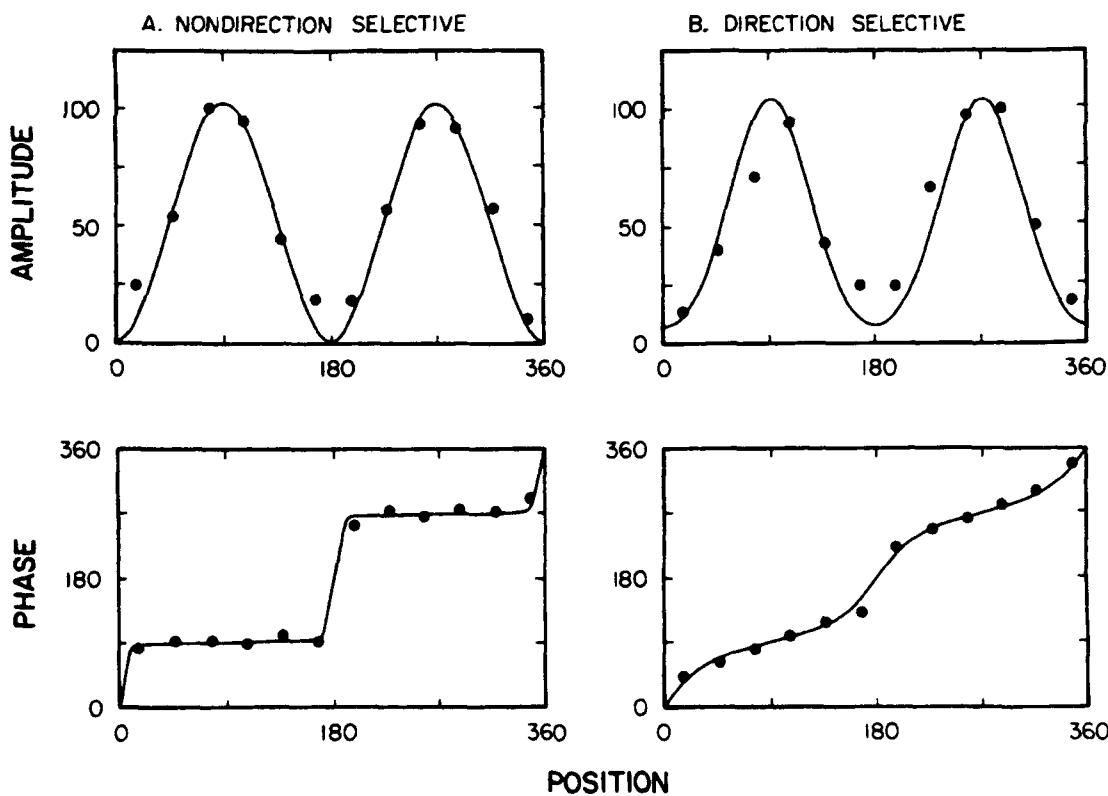


**Figure 7.** Amplitude and phase responses of a nondirection selective cell (A) and a direction selective cell (B) to stationary gratings counterphase flickering in different spatial positions. The smooth curves show what would be expected from a linear filter.<sup>14</sup>

Figure 7 plots the expected and the measured responses of a direction selective cell and a nondirection selective cell to stationary counterphase flickering gratings, presented in different spatial positions. The panels on the left show the amplitude and phase of response for a nondirection selective simple cell as a function of the position of a counterphase flickering grating. From the work of Enroth-Cugell and Robson<sup>23</sup> as well as Hochstein and Shapley,<sup>24</sup> we know that the response should be a sinusoidal function of the spatial position of the grating, with two null phase positions (that is, two spatial phase positions which evoke little or no response); the smooth lines through the data points show the predictions of a strictly linear filter. For the nondirection selective cell on the left, the fit is reasonable.

Now consider the predicted and measured responses for the direction selective cell, on the right. This particular cell produced almost no response to gratings drifting in the nonpreferred direction -- it was very direction selective. Given a strictly linear cell with this degree of direction selectivity, the amplitude of response to a counterphase flickering grating would not change with spatial position and the phase would change continuously. This is because, a counterphase flickering grating can be decomposed into two gratings of equal contrast drifting in opposite directions. A strictly linear direction selective filter would only be affected by the component drifting in the preferred direction. The amplitude of this component remains constant -- and the phase changes continuously. These predictions are illustrated by the solid lines. As can be seen, this direction selective simple cell does not behave according to these strictly linear predictions.

We have shown that this kind of behavior can be readily accounted for if the expansive exponent of the contrast response function is taken into consideration. In Figure 8, the same responses are plotted along with the predictions of a model composed of a linear filter followed by the measured contrast response exponent. The fit is good for both the direction selective cell and the nondirection selective cell.

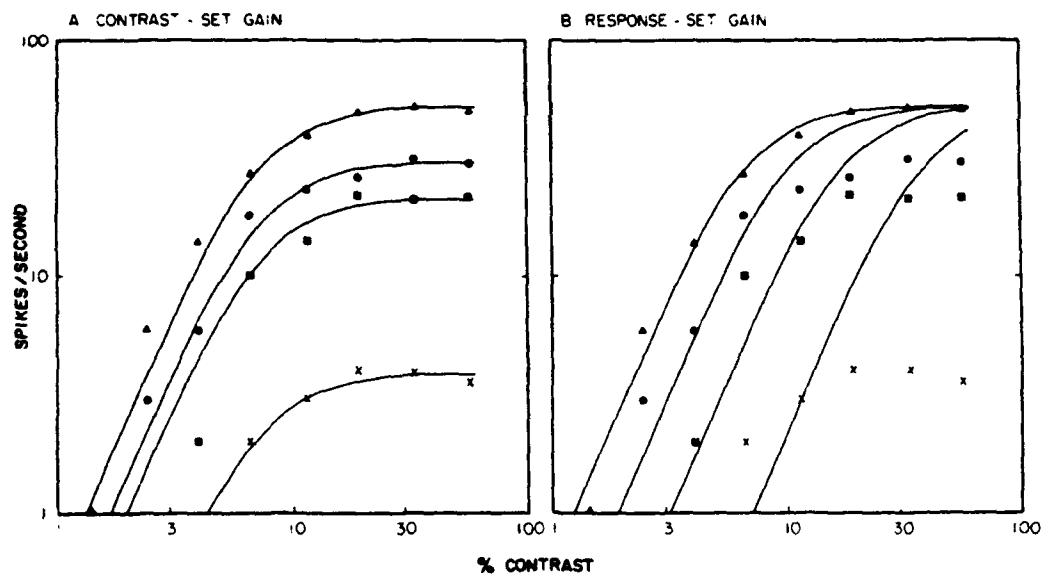


**Figure 8.** Amplitude and phase responses of a nondirection selective cell (A) and a direction selective cell (B) to stationary gratings counterphase flickering in different spatial positions. The smooth curves show the predictions from a model composed of a linear filter followed by the measured exponent of the contrast response function.<sup>14</sup>

## 5. CONTRAST GAIN MAINTAINS SELECTIVITY

The steep slopes of the contrast response function force most cortical cells to have a limited dynamic response range. As a function of contrast, the response increases rapidly and then saturates. While the steep slopes of the contrast response function may well enhance stimulus selectivity, the saturation nonlinearity could potentially have very deleterious effects on the overall stimulus selectivity of the cortical cells.

Consider, for example, what would happen to spatial frequency selectivity if the saturation were due to a limitation imposed by the final response generating mechanism of the cortical cell, after summation of inputs. Under these circumstances, the cell would exhibit very narrow spatial frequency tuning when measured at low contrasts, but then exhibit very broad spatial frequency tuning when measured at high contrasts. This is what we expected to find. The validity of these expectations can be tested by measuring the spatial frequency tuning at multiple contrasts, or equivalently, by measuring the contrast response function across a range of spatial frequencies -- particularly, optimal and nonoptimal spatial frequencies.

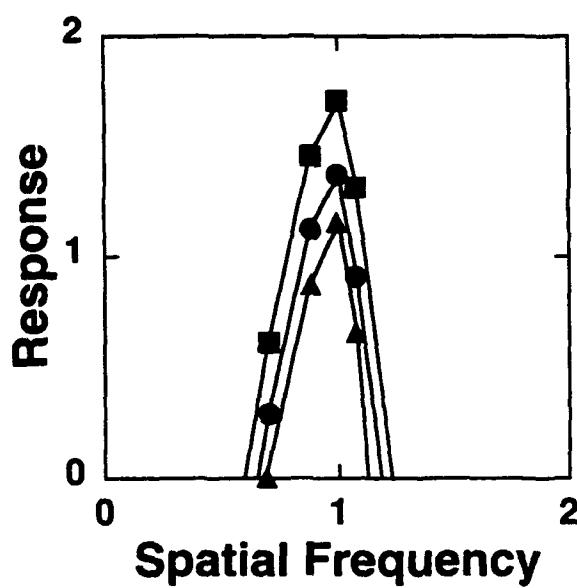


**Figure 9.** Contrast response functions measured at different spatial frequencies. The panel on the left (**A**) shows the predictions for a *contrast-set gain* model; the panel on the right (**B**) shows the predictions for a *response-set gain* model.<sup>10</sup>

Figure 9 illustrates this point. The solid lines in the panel on the right are the predicted contrast response functions, measured at different spatial frequencies, given a saturation that is dependent upon the overall level of the final response of the cell. On log coordinates, the curves shift horizontally, the maximum response rate stays the same, only the semi-saturation constant changes. This is the response-set gain model; it is what

we were expecting to find. At a low contrast, say 5%, an optimal spatial frequency produces some 50 spikes/second whereas a nonoptimal frequency produces only 1 spike/second -- a 50-to-1 difference; the cell is very selective. However, at a high contrast, say 50%, the selectivity is gone since both the nonoptimal and the optimal evoke the same maximum saturated response. This is not how cortical cells behave; these predicted curves do not fit the measured responses.

The panel on the left shows the same measured responses along with the predictions of a different model. In this case, the saturation is not determined by the overall level of the response but rather the gain is set by the overall level of the contrast. On log coordinates, the curves shift vertically, the semi-saturation constant remains the same and only the maximum response changes. This is the *contrast-set gain model* of Albrecht and Hamilton,<sup>10</sup> which we will shorten here to *contrast gain model*. As can be seen, the contrast gain model provides a much better fit to the measured responses. Saturation tends to occur at the same contrast level for all the different spatial frequencies, not at the same response level. Further, the magnitude of the saturated response is different for each spatial frequency. Thus, the relative response ratios between spatial frequencies are maintained across contrast.



**Figure 10.** Spatial frequency tuning functions measured at three different contrasts.<sup>10</sup>

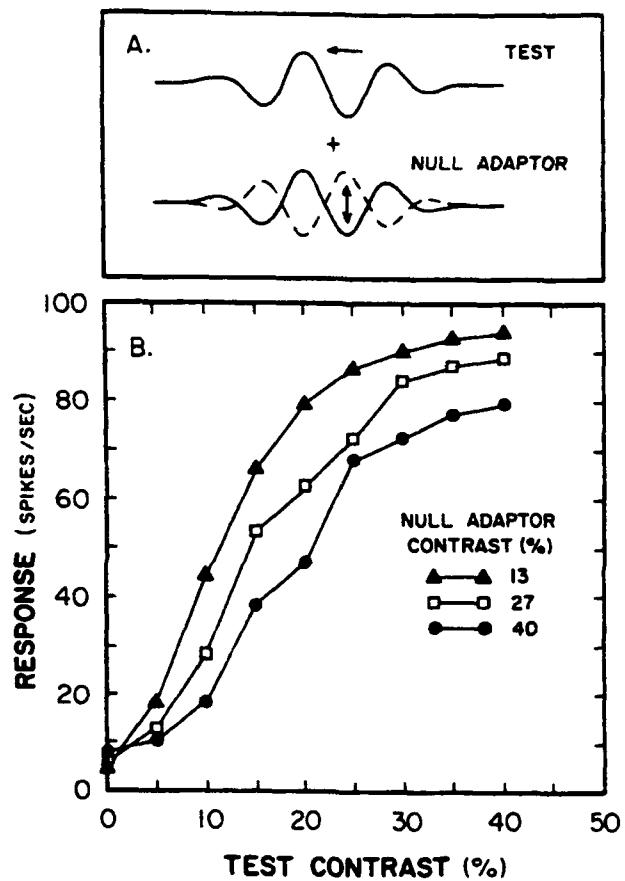
One overall net effect of this contrast gain mechanism is to preserve the spatial frequency selectivity: the relative response ratios between spatial frequencies are maintained across contrast -- the responses of nonoptimal spatial frequencies remain non-optimal even at high contrasts. Figure 10 shows the spatial frequency tuning of a typical cell measured at different contrasts. As can be seen, the overall shape and bandwidth change very little as the contrast of the gratings is varied from a threshold value of 4% through a midrange value of 6.6% and a saturation value of 33%. The tuning remains relatively invariant with contrast. The bandwidth remains nearly fixed at 0.7 octaves. Over the last decade, many different laboratories have replicated and extended this basic finding to all of the important dimensions of stimulus selectivity: orientation selectivity, direction selectivity, ocular selectivity, spatial phase selectivity, etc.<sup>10,11,14,25,26</sup>

## 6. ISOLATION OF CONTRAST GAIN CONTROL

We have just begun to explore the spatial and temporal properties of the contrast gain control mechanism using a new technique which we call the *null-adaptor technique*.<sup>16</sup> To isolate contrast gain control we vary the contrast of a counterphase grating which is (a) located at the null phase position and (b) confined in length and width to the conventional receptive field. This stimulus does not evoke a response from the cell but it does allow us to control the average contrast while holding other factors constant (such as response fatigue, slow adaptation, spatial frequency and orientation inhibition, etc.). To assess the effect of the adapting contrasts, we superimpose a drifting grating of the same spatial frequency, temporal frequency, orientation, length, and width. Figure 11A illustrates this basic stimulus configuration.

Figure 11B plots the contrast response function (measured with the drifting grating) in the presence of three adapting contrasts. As can be seen, the three null adaptor contrasts had little or no effect on the response of the cell when presented alone (i.e. when the contrast of the drifting test was zero); however, the adapting contrasts had a substantial effect on the responses to the superimposed drifting grating. Specifically, as the adapting contrast increased, sensitivity to the drifting contrast decreased; the contrast response function primarily shifts to the right.

These curves indicate that while the stationary flickering grating evokes no response from the cell in the null phase position, it nevertheless controls the overall sensitivity of the cell to contrast through a fast-acting, multiplicative, gain control mechanism.



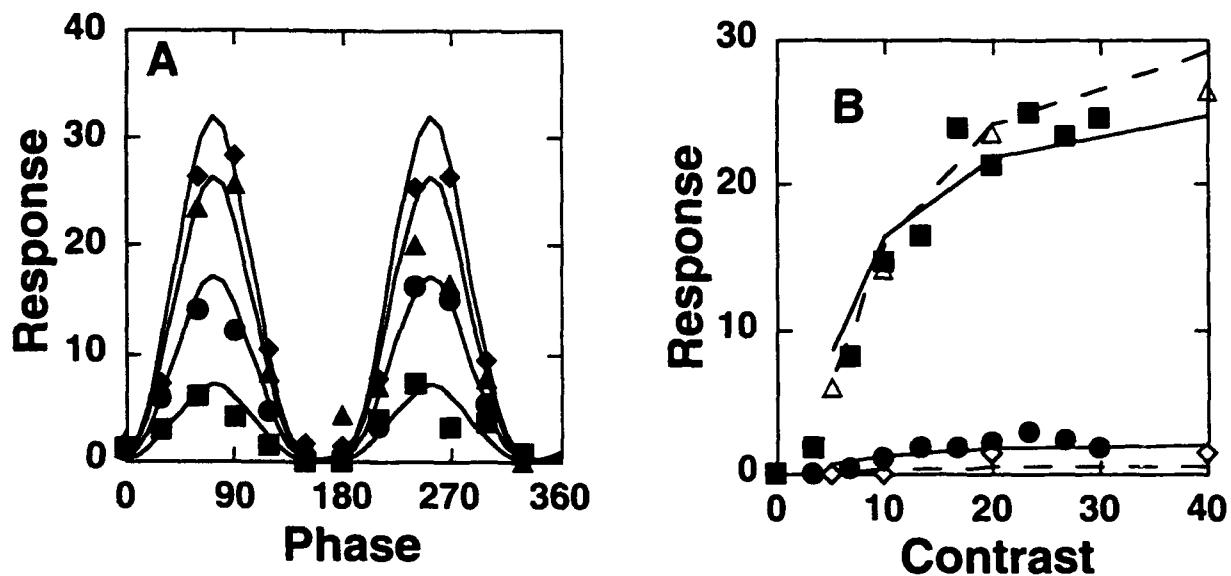
**Figure 11.** (A) Stimulus configuration used to isolate contrast gain; a drifting grating is superimposed upon a counterphase grating flickering at the "null phase position." (B) Contrast response functions measured with a drifting grating in the presence of different null adaptor contrasts.<sup>16</sup>

## 7. APPENDIX: STATIONARY & DRIFTING GRATING AMPLITUDES

Over the past few years, three different laboratories have performed a similar set of experiments that were designed to investigate the fundamental mechanism responsible for direction selectivity.<sup>14,27,28</sup> In these studies, the responses to drifting sine wave gratings were compared with the responses to stationary counterphase flickering gratings. All three reports demonstrated that the results were not totally consistent with what would be expected based upon simple linear summation over a receptive field oriented in the space-time domain; for example, the direction selectivity predicted from responses to flickering gratings generally underestimated the direction selectivity measured from the responses to drifting gratings.

The results contained within the three reports were quite similar; further, all three reports seemed to agree that while linear summation could probably account from some of the direction selectivity, an additional nonlinear contribution would be required to account for the degree of direction selectivity. Reid et al. proposed a model in which the direction selectivity of a linear filter was "sharpened" by nonlinear suppression of the responses in the nonpreferred direction. Tolhurst and Dean proposed a similar model. As described above, Albrecht and Geisler incorporated the nonlinearities evident in the contrast response function (contrast gain control and expansive response exponent) and found that the discrepancies between the measured and predicted responses, to drifting and flickering gratings, were diminished.

Reid et al., and Tolhurst and Dean, compared the absolute magnitude of the response to drifting gratings with a simple linear prediction based upon the measured responses to flickering gratings. The linear predictions are straight forward; the response in the preferred direction of motion should be equal to the sum of the peak and the trough from the counterphase data while the response in the nonpreferred direction should be equal to the difference of the peak and the trough. They found that, for the preferred direction of motion, the measured responses were approximately equal to the linear predictions. However, for the nonpreferred direction of motion, the measured responses were considerably less than the linear predictions. This result may be consistent with what might be expected from a "nonlinear direction-selective suppression mechanism." However, this simple linear prediction ignores the well-known nonlinearities evident in the contrast response function (saturation due to contrast gain and expansive response exponents). As Heeger<sup>18</sup> recently pointed out, this pattern of results is what might be expected when these two nonlinearities are taken into consideration.



**Figure 12.** Responses to stationary and drifting gratings along with a model composed of a linear filter, contrast gain control and expansive response exponent. (A) Responses to counterphase gratings in different positions and contrasts; the smooth curves show the fit of the contrast-gain/exponent model. (B) Contrast response function measured with drifting gratings in the preferred direction of motion (filled squares) and nonpreferred direction (filled circles); the smooth curves (solid lines) show the *predictions* of the contrast-gain/expONENT model based upon the *fit* to the counterphase responses; open triangles and the dashed line replot the counterphase responses and fit near the peak; open diamonds and the dashed line replot the counterphase responses and fit near the trough.

Figure 12A plots the responses of a direction selective simple cell (recorded from the striate cortex of a macaque monkey) to a counterphase grating flickering in different spatial positions at four separate contrasts. The smooth curves show the fit of a model which incorporates the nonlinearities evident in the contrast response function; specifically, the contrast-gain/expONENT model (formally described elsewhere).<sup>14</sup> Given strict linearity, the null phase positions would lead to the erroneous conclusion that this cell was nondirection selective and that the responses to gratings drifting in either direction would be equal to the responses at the optimal position of the counterphase gratings. (The sum and the difference of the peak and the trough are obviously equal when the trough is zero.) Figure 12B plots the responses of the same cell to gratings drifting in the preferred and nonpreferred direction as a function of contrast; the responses to the counterphase grating (from 12A, near the peak and trough position), are superimposed. As can be seen, while the responses to the grating drifting in the preferred direction are approximately equal to the linear prediction (i.e., the responses are approximately the sum of the peak and the trough), the responses in the nonpreferred direction are far below the linear prediction (i.e., the responses are far below the difference of the peak and the trough). The smooth curves are the predictions of the

contrast-gain/exponent model using the optimized parameters from the counterphase data. As can be seen, the model conforms well to the measured responses.

Reid et al., and Tolhurst and Dean, summarize their data for the total sample of cells using scatter plots (one for each direction of motion), where the x-axis is the measured response to drifting gratings, and the y-axis is the linear prediction from the measured responses to counterphase gratings. We have performed a similar analysis on both a sample of cat and a sample of monkey striate cortex neurons; the results are very similar to what Reid et al. and Tolhurst and Dean reported. In general (across all laboratories, in both cat and monkey), for the preferred direction of motion, the data cluster around the diagonal (congruent with the linear predictions); however, for the nonpreferred direction of motion, the data cluster above the diagonal (contrary to the linear predictions).

Further, two clear differences are evident in a comparison of the scatter plot for the preferred direction of motion with the scatter plot for the nonpreferred direction: the nonpreferred data points are more dispersed and the regression line is shifted toward the upper left whereas the preferred data points are less dispersed and the regression line is shifted slightly toward the lower right corner. This pattern of results is consistent with what would be expected of a random sample of visual cortex neurons having contrast gain control, the known distribution of direction selectivities and the known distribution of response exponents. As summarized in Table 1, the location of the regression line and the degree of dispersion can be affected by: the contrast gain control, the degree of direction selectivity, and the value of the expansive exponent.

For the shift in the regression line, the arrows in Table 1 summarize the following relationships: (a) as the exponent increases from 1.0 (given any degree of direction selectivity), there is an asymmetric shift in the flicker predictions to overestimate the responses in the nonpreferred direction and underestimate the responses in the preferred direction; that is, the regression line shifts toward the upper left corner for nonpreferred and the lower right corner for preferred; (b) as the direction selectivity increases (given an expansive exponent), there is a similar asymmetric shift in the flicker predictions; that is, the regression line shifts toward the upper left corner for the nonpreferred direction and the lower right corner for the preferred direction; (c) as the contrast gain control factor is increased (given the difference in the spatiotemporal RMS contrast of a counterphase grating vs. a drifting grating -- equated using the conventional peak to trough "Michelson" contrast), the regression line shifts toward the upper left corner for both directions.

**Table 1:** Effects of the contrast gain control (Gain), the direction selectivity (Dir), and the expansive response exponent (Exp) on the shift of the scatter plot regression line and degree of dispersion for the preferred and nonpreferred directions of motion. Direction of the arrow indicates the direction of the effect; open arrows indicate that the effect is minor.

		Gain	Dir	Exp
Shift	Non	↑	↑	↑
	Pref	↑	↓	↓
Dispersion	Non	↑	↑	↑
	Pref	↓	↑	↑

For the degree of dispersion, the arrows in Table 1 summarize the following relationships: (a) as the exponent increases from 1.0 (given some level of variability in the degree of direction selectivity from cell to cell) there is a greater degree of dispersion for the nonpreferred direction of motion as opposed to the preferred direction of motion; (b) as the degree of direction selectivity is increased (given some level of variability in the exponent from cell to cell), there is a greater degree of dispersion for the nonpreferred direction of motion as opposed to the preferred direction of motion; (c) as the contrast gain factor is increased (given some level of variability in either/both the degree of direction selectivity or/and the exponent), the degree of dispersion increases for the nonpreferred but decreases for the preferred.

In summary, the data in the scatter plots reveal a consistent pattern: asymmetric shift of the regression line and the degree of dispersion, depending upon drift direction. This pattern of results is consistent with what one might expect given a random sample of cortical cells and the known effects of the nonlinearities seen in the contrast response function: the cell to cell variation in the degree of direction selectivity, along with the cell to cell variation in the value of the exponent, would combine with the differential contrast gain to produce the asymmetric shift and the asymmetric dispersion.

## 8. ACKNOWLEDGMENTS

This research was supported by TARP grant 003658-463, by AFOSR grant F49620-93-1-0307, by the Primate Vision Endowment (DGA), and by NIH grant EY02688 (WSG). The authors thank Larry Stern for his assistance in all phases of this research.

## 9. REFERENCES

1. Robson, J.G., "Receptive fields: neural representation of the spatial and intensive attributes of the visual image." In, *Handbook of Perception, Volume V*, E.C. Carterette and M.P. Friedman, eds. New York: Academic Press. 1975.
2. Robson, J.G., "Frequency domain visual processing." In, *Physical and Biological Processing of Images*, O.J. Braddick and A.C. Sleigh, eds. New York: Springer-Verlag. 1983.
3. De Valois, R.L. and K.K. De Valois, *Spatial Vision*. New York: Oxford University Press. 1988.
4. Movshon, J.A., I.D. Thompson, and D.J. Tolhurst, "Spatial summation in the receptive fields of simple cells in the cat's striate cortex." *Journal of Physiology, London*, **383**: 53-77. 1978.
5. De Valois, R.L., D.G. Albrecht, and L.G. Thorell, "Spatial frequency selectivity of cells in macaque visual cortex." *Vision Research*, **22**: 545-559. 1982.
6. Jones, J.P. and L.A. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex." *Journal of Neurophysiology*, **58**: 1233-1258. 1987.
7. Hamilton, D.B., D.G. Albrecht, and W.S. Geisler, "Visual cortical receptive fields in monkey and cat: Spatial and temporal phase transfer function." *Vision Research*, **29**: 1285-1308. 1989.
8. McLean, J. and L. Palmer, "Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat." *Vision Research*, **29**: 675-679. 1989.
9. DeAngelis, G.C., I. Ohzawa, and R.D. Freeman, "Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation." *Journal of Neurophysiology*, **69**: 1118-1135. 1993.
10. Albrecht, D.G. and D.B. Hamilton, "Striate cortex of monkey and cat: Contrast response function." *Journal of Neurophysiology*, **48**: 217-237. 1982.
11. Li, C. and O. Creutzfeldt, "The representation of contrast and other stimulus parameters by single neurons in area 17 of the cat." *Pflugers Archive*, **401**: 304-314. 1984.

12. Sclar, G., J.H.R. Maunsell, and P. Lennie, "Coding of image contrast in central visual pathways of macaque monkey." *Vision Research*, **30**: 1-10. 1990.
13. Bonds, A.B., "Temporal dynamics of contrast gain in single cells of the cat striate cortex." *Visual Neuroscience*, **6**: 239-255. 1991.
14. Albrecht, D.G. and W.S. Geisler, "Motion selectivity and the contrast-response function of simple cells in the visual cortex." *Visual Neuroscience*, **7**: 531-546. 1991.
15. Robson, J.G., "Neural coding of contrast in the visual system." *Optical Society of America, Technical Digest Series*, **17**: 152. 1991.
16. Geisler, W.S. and D.G. Albrecht, "Cortical neurons: Isolation of contrast gain control." *Vision Research*, **32**: 1409-1410. 1992.
17. Heeger, D.J., "Computational model of cat striate physiology." In, *Computational Models of Visual Perception*, J.A. Movshon and M. Landy, eds. Cambridge: MIT Press. 1991.
18. Heeger, D.J., "Modeling simple cell direction selectivity with normalized, half-squared, linear operators." *Investigative Ophthalmology & Visual Science Annual Meeting*, **33**: 953. 1992.
19. Heeger, D.J., "Normalization of cell responses in cat striate cortex." *Visual Neuroscience*, **9**: 191-197. 1992.
20. Heeger, D.J., "Half-squaring in responses of cat striate cells." *Visual Neuroscience*, **9**: 427-443. 1992.
21. Reid, R.C., R.E. Soodak, and R.M. Shapley, "Linear mechanisms of directional selectivity in simple cells of cat striate cortex." *Proceedings of the National Academy of Sciences of the U.S.A.*, **84**: 8740-8744. 1987.
22. Hamilton, D.B., *The phase transfer function of visual cortical neurons*. Austin: University of Texas. 1987.
23. Enroth-Cugell, C. and J.G. Robson, "The contrast sensitivity of retinal ganglion cells of the cat." *Journal of Physiology*, **187**: 517-552. 1966.
24. Hochstein, S. and R.M. Shapley, "Quantitative analysis of retinal ganglion cell classifications." *Journal of Physiology*, **262**: 237-264. 1976.
25. Sclar, G. and R.D. Freeman, "Orientation selectivity in the cat's striate cortex is invariant with stimulus contrast." *Experimental Brain Research*, **46**: 457-461. 1982.
26. Skottun, B.C., A. Bradley, G. Sclar, I. Ohzawa, and R.D. Freeman, "The effects of contrast on visual orientation and spatial frequency discrimination: A comparison of single cells and behavior." *Journal of Neurophysiology*, **57**: 773-786. 1987.
27. Reid, R.C., R.E. Soodak, and R.M. Shapley, "Directional selectivity and spatiotemporal structure of receptive fields of simple cells in cat striate cortex." *Journal of Neurophysiology*, **66**: 505-529. 1991.
28. Tolhurst, D.J. and A.F. Dean, "Evaluation of a linear model of directional selectivity in simple cells of the cat's striate cortex." *Visual Neuroscience*, **6**: 421-428. 1991.

A New Model of Human Luminance Pattern Vision Mechanisms: Analysis of the Effects of Pattern Orientation,  
Spatial Phase and Temporal Frequency

John M. Foley and Geoffrey M. Boynton  
University of California, Santa Barbara  
Department of Psychology  
Santa Barbara, California 93106

ABSTRACT

Models of human pattern vision mechanisms are examined in light of new results in psychophysics and single-cell recording. Four experiments on simultaneous masking of Gabor patterns by sinewave gratings are described. In these experiments target contrast thresholds are measured as functions of masker contrast, orientation, spatial phase and temporal frequency. The results are used to test the theory of simultaneous masking proposed by Legge and Foley that is based on mechanisms that sum excitation linearly over a receptive field and produce a response that is an S-shaped transform of this sum. The theory is shown to be inadequate. Recent single-cell recording results from simple cells in the cat show that these cells receive a broadband divisive input as well as an input that is summed linearly over their receptive fields. A new theory of simultaneous masking based on mechanisms with similar properties is shown to describe the psychophysical results well. Target threshold vs masker contrast (TvC) functions for a set of target-masker pairs are used to estimate the parameters of the theory including the excitatory and inhibitory sensitivities of the mechanisms along the various pattern dimensions. The human luminance pattern vision mechanisms, unlike most of the cells, do not saturate at high contrast.

1. INTRODUCTION

Neurobiology and human psychophysics have had a long interaction which has been particularly fruitful in the field of pattern vision. Mach bands and related phenomena suggested the existence of lateral connections in the visual system long before they were shown biologically. Single-cell recording established that cells are tuned on to spatial frequency and that their contrast-response functions are nonlinear. These findings provided the basis for a theory of simultaneous masking<sup>1</sup> the essence of which was that one pattern will mask another when the masking pattern excites the same mechanisms that detect the target. It was proposed that masker excitation produces masking because of the compressive nonlinearity in the response function; a target on top of a masker adds a smaller increment to the response than a target alone. Superimposing a low contrast pattern on a target was found to decrease the threshold for the target (facilitation). This psychophysical finding suggested an accelerating nonlinearity at low contrast<sup>1,2</sup>. Such a nonlinearity was later found in single unit studies<sup>3</sup>.

The Legge and Foley theory of masking was based on mechanisms that sum excitation linearly over an orientated receptive field of medium bandwidth and then produce a response that is an S-shaped function of that sum. This idea has come to be widely accepted, so much so that it was recently referred to as the "standard model" of masking<sup>4</sup>. Unfortunately, it was never adequately tested, although psychophysical and biological evidence for very broadband interaction in pattern vision has existed for some time<sup>5</sup>. The clearest and most complete evidence comes from single unit recording of cortical cells in the cat done by Bonds<sup>6</sup>. Bonds suggested that this interaction is divisive and Heeger<sup>7</sup> has recently proposed a model of cat cortical cell responses in which a broadband inhibitory signal is divisive of the response to excitation.

These developments led us first to test the Legge and Foley theory of simultaneous masking and finding it to fail, to propose a new theory of that was inspired by the biological developments. This new theory fits psychophysical data well and is largely consistent with the single cell work. It leads to some counterintuitive predictions that have been borne out by psychophysical experiments. In this paper we describe the theory and some of the experiments that have been done to test it and to pursue its implications.

## 2. A NEW MODEL OF HUMAN PATTERN VISION MECHANISMS

Figure 1 illustrates the model of the human pattern vision mechanisms that is the basis of our new theory of simultaneous pattern masking. Here the response of the mechanisms depends not only on the net excitation of the receptive field by the pattern, but also on inhibitory inputs which are more broadly tuned to pattern features. The theory is formulated so that any pattern may be described as a sum of component patterns, and there is an excitatory term and an inhibitory term for each pattern component. In our experiments the components are sinewave gratings and Gabor patterns of different orientations, spatial phases, and spatial and temporal frequencies. Excitation sums linearly across components before being halfwave rectified and raised to the power p. Inhibition sums linearly only for similar components and these partial sums are each raised to the power q before being summed together to produce the inhibitory term in the denominator. The Z in the denominator is a constant in our experiments, but it may depend on the past stimulation of the mechanism. In fitting the model to data we have found that  $p > q$  and  $p, q > 2$ .

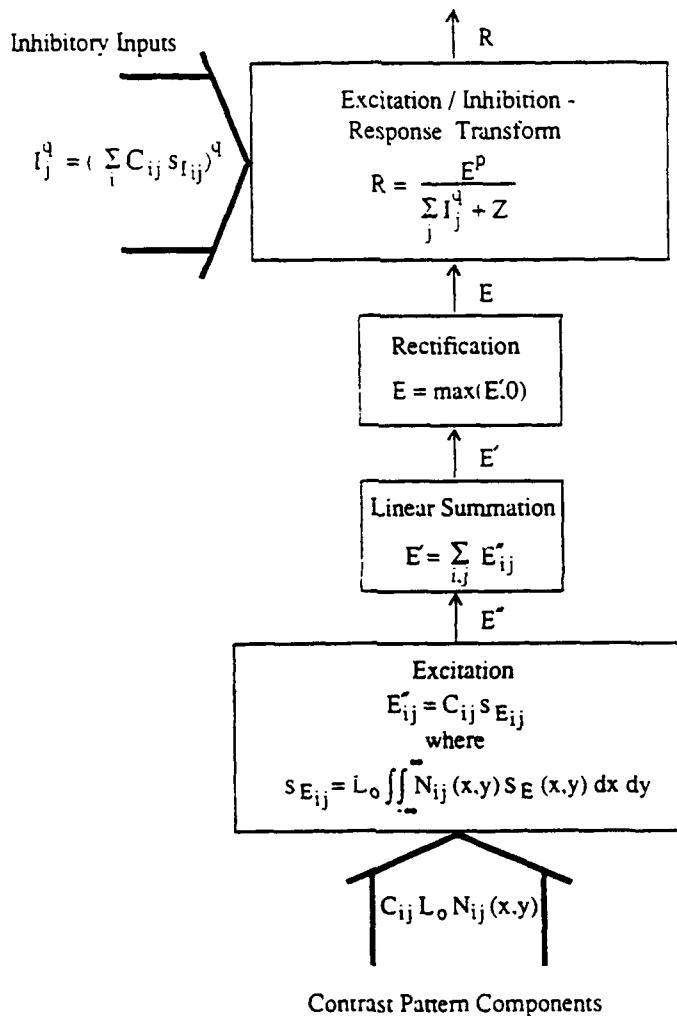


Figure 1. The new model of the human luminance pattern vision mechanisms. In addition to linear summation of excitation over a receptive field the mechanism receives a broadband inhibitory input which is divisive of the excitation.

It is of interest to look at the response of these mechanisms to some simple stimuli. We have assigned parameters to the theory that are similar to those that we need to account for psychophysical data. On the left of Figure 2 are response functions for five levels of excitatory sensitivity when inhibitory sensitivity is held constant. The functions are S-shaped and are attenuated by a multiplicative factor as excitatory sensitivity decreases as, for example, it does when stimulus orientation changes from optimal to increasingly nonoptimal values. On the right are shown the same functions in log-log coordinates. These look a lot like contrast-response functions of simple cortical cells and they change in a similar way when the stimulus becomes less optimal<sup>3</sup>. One difference is that cells generally saturate within this range, but these human pattern mechanism responses do not. Actual functions inferred from psychophysical data are displaced downward more at high contrast than at low as stimulus orientation changes. This indicates that both excitatory and inhibitory sensitivity decrease as the stimulus becomes increasingly different than the optimal stimulus.

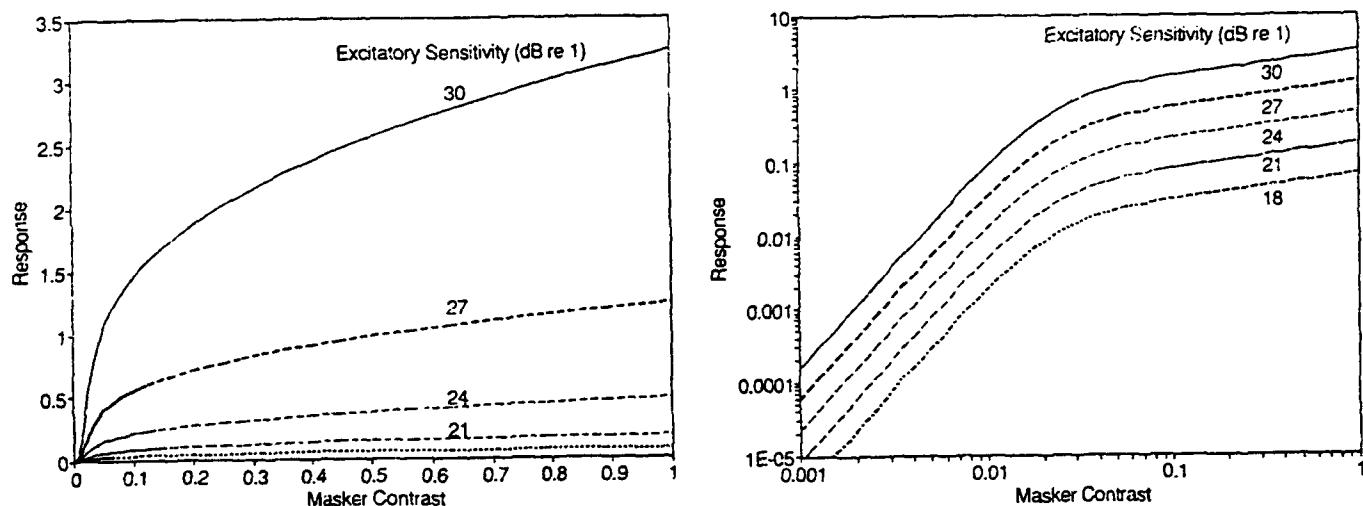


Figure 2. Contrast response functions for human pattern vision mechanisms with excitatory sensitivity as a parameter. Inhibitory sensitivity is held constant. The response decrease approximates what happens when the stimulus is rotated away from the optimal orientation. a) in linear coordinates. b) in log-log coordinates.

Like the old theory, the new theory assumes that when a target is superimposed on top of a masker, it will be at threshold when the response to the target plus masker minus the response to the masker alone (the target increment) equals 1. Figure 3 shows the response functions for a case when the excitatory sensitivity to the masker is high and one when the excitatory sensitivity to the masker is 0. Here the target contrast is fixed at 0.1. The masker has radically different effects on the mechanism response in the two cases, but the target increment decreases with masker contrast in a similar way in both cases. In the first case the threshold depends on both excitation and inhibition produced by the masker. In the second, the masker produces essentially no excitation, but as masker contrast increases, masker produced inhibition reduces the response to the target. Thus, response compression due to masker excitation of the detecting mechanism is not the process that produces masking, and the relation between excitatory sensitivity and masking is more complex than we previously thought.

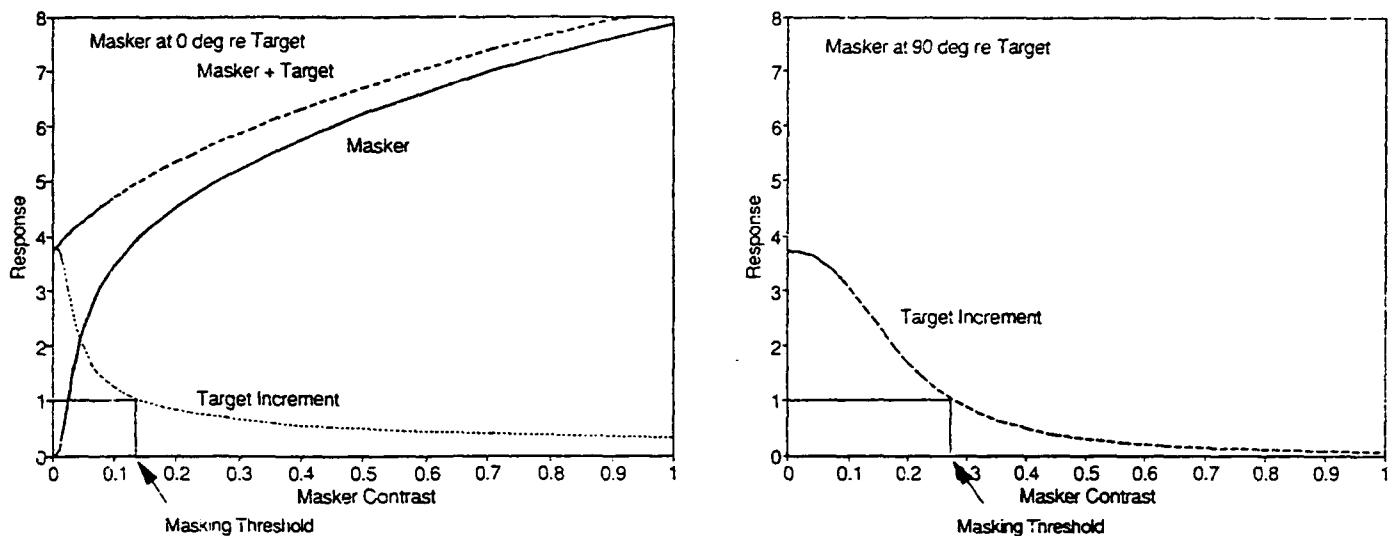


Figure 3. Response to a masker, masker plus target of contrast = 0.1, and their different (target increment). Parameters are based on the data of experiment 1. a) Excitatory sensitivity to masker is high. b) Excitatory sensitivity to masker is 0.

### 3. EXPERIMENTS

We will describe four simultaneous masking experiments that test this new theory and estimate its parameters. In the first three experiments the methods were very similar. The targets were symmetric vertical Gabor patterns whose 1/e halfwidth corresponded to 1 period of the grating. They were in cosine phase with the fixation point and centered on it. The maskers were sinewave gratings that filled the field of 5 deg high by 7 deg wide. Target and masker were presented simultaneously as 33 msec pulses. Background luminance was 32 cd/m<sup>2</sup> and viewing distance was 162 cm. A two-alternative temporal forced-choice paradigm and an adaptive threshold-seeking algorithm were used<sup>8</sup>.

Most of the experiments measure the target contrast threshold as a function of masker contrast (TVC function). In each experiment this is done for a set of target and masker pairs in which the target waveform is constant and the masker waveform varies along some dimension. The Legge and Foley theory makes a strong prediction about these functions. The theory implies that, when a single mechanism or a set of mechanisms with very similar contrast sensitivity mediate detection any change in the spatial waveform of the masker will shift this function along the masker contrast axis by a multiplicative constant (multiplicative horizontal displacement). It will be seen that this prediction fails repeatedly in these experiments, but the new theory describes the results well.

### 3.1 Exp. 1: Effect of masker orientation on TvC function

Experiment 1 tested for multiplicative horizontal displacement over changes in the orientation of the masker<sup>9</sup>. Both the target center spatial frequency and the masker frequency were 2 c/deg. The target was vertical and the masker varied over an orientation range of 0-90 deg re vertical. Both target and masker were in cosine phase with the fixation point.

The TvC functions for one observer are shown in Figure 4. Contrast is defined as (peak luminance - background luminance)/background luminance. Contrast is expressed in decibels re 1 where 1 dB corresponds to 1/20 of a log unit ( $C_{dB} = 20 \times \log_{10} C$ ). The standard deviation of these measurements is approximately 1 dB at low contrasts and increases to about 2 dB at high contrasts. Considerable masking occurs at all relative orientations including 90 deg. Note that threshold elevation is greater at 22.5 deg than at 0 deg. Facilitation, on the other hand, occurs only for small orientation differences. Within the masking range, there is also variation in the form of the TvC functions. In this range they are concave downward when the orientation difference is 0 and they become more linear and shallower as orientation difference increases. Thus, multiplicative horizontal displacement fails for the system when target-masker orientation difference is varied over a range of 0-90 deg. The smooth curves correspond to the best fit of the new theory to the data.

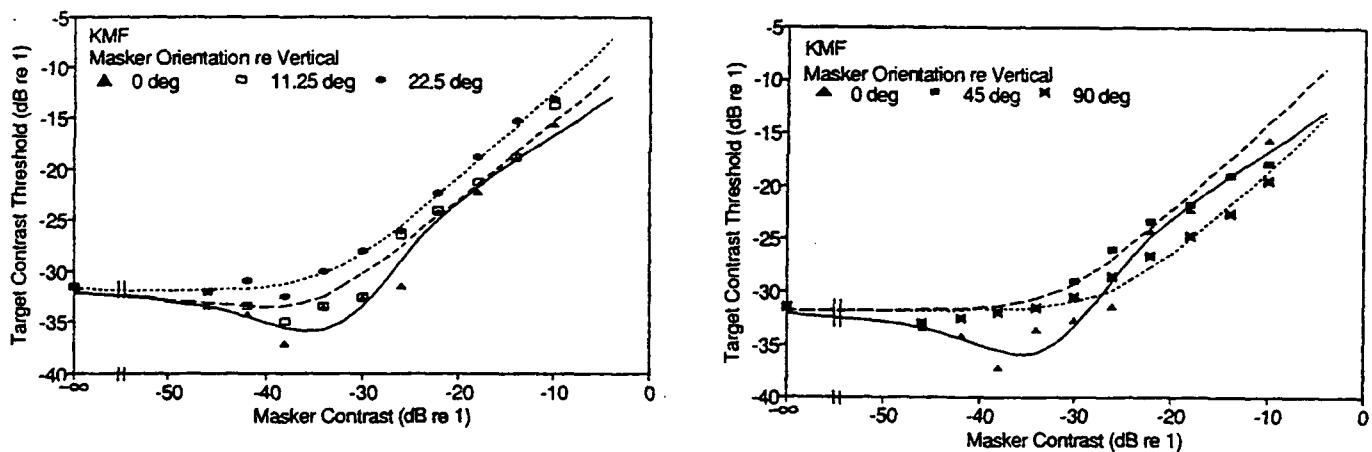


Figure 4. Experiment 1. TvC functions for simultaneous masking with masker orientation re target as the parameter. The function for relative orientation = 0 deg is in both panels. Smooth curves correspond to the best fitting version of the new theory.

It is possible that different mechanisms with different parameters mediate detection at different masker orientations. There are two kinds of evidence against this. First, if the detecting mechanism changes, it is reasonable to assume that the percept will change. I tested for a change in the percept by having the observer give phenomenological descriptions of the appearance of the stimulus when the target was at threshold. These were made immediately after each threshold measurement. Although there was variability in these reports, there was no change related to masker orientation. For all masker orientations, when an orientation was associated with target presence, it was almost always vertical. Second, at absolute threshold the same mechanism is likely to detect in all

conditions because the stimulus is the same. If other mechanisms intrude as masker contrast increases, a characteristic scallop would be expected in the Tvc function<sup>10</sup>. This is not seen in these data. Consequently, we assume that the same mechanism or a set of similarly tuned mechanisms mediated detection in all conditions. The results show that the horizontal displacement rule fails for these mechanisms.

Figure 5 shows the excitatory and inhibitory sensitivities to the masker as a function of orientation. These sensitivities are parameters of the new theory that were estimated by fitting the theory to the data of experiment 1. Excitatory sensitivity falls off rapidly with orientation and is essentially 0 by 22.5 deg. Inhibitory sensitivity falls off much more gradually and is still substantial at 90 deg. It may have a slight maximum around 11 deg.

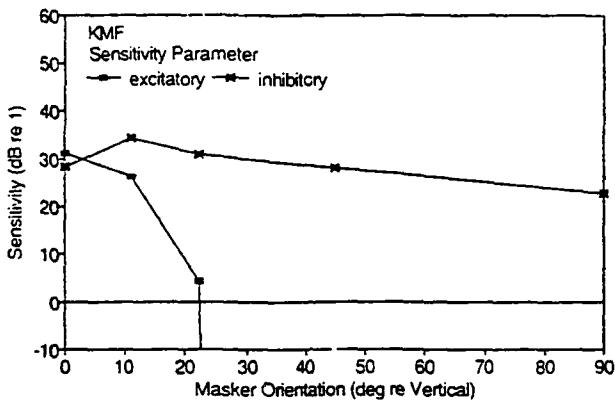


Figure 5. Excitatory sensitivity and inhibitory sensitivity as a function of masker orientation re target. Values were estimated by fitting the new theory to the data of experiments 1 and 2.

### 3.2 Exp. 2: Effect of masker spatial phase

Experiment 2 is a first look at the effect of masker spatial phase relative to the target. Masker spatial phase is varied over a range of +/- 90 deg re the target. Here the Legge and Foley theory predicts that there will be some relative phase within this range at which masking will go to zero. This follows because the convolution of a sinewave with any function will yield a sinewave of the same frequency and any sinewave will pass through 0 in any 180 deg range. The results are shown in Figure 6. The horizontal line corresponds to the absolute threshold. There is substantial masking at each of the relative phases. Unless the function takes a very odd form between the data points, masking does not go to zero and the results are inconsistent with the old theory. More interesting at this point is the finding that masking increases with increasing phase difference over this range.

### 3.3 Exp. 3: Effect of masker spatial phase on Tvc function

Experiment 3 measured Tvc functions for maskers at three spatial phases of the masker relative to the target: -90, 0 and +90 deg. Targets were again vertical Gabor patterns and maskers were vertical sinewave gratings of the same frequency. The results are shown in Figure 7. At 0 deg the familiar dipper-shaped function was obtained. At 90 deg, no facilitation was found and masking was greater than at 0 deg throughout the masker

contrast range. The smooth curves through these data points correspond to a version of the theory in which the inhibitory sensitivity of the mechanisms is independent of the spatial phase of a stimulus grating, but excitatory sensitivity, which is positive at 0 deg, goes to 0 at 90 deg relative phase.

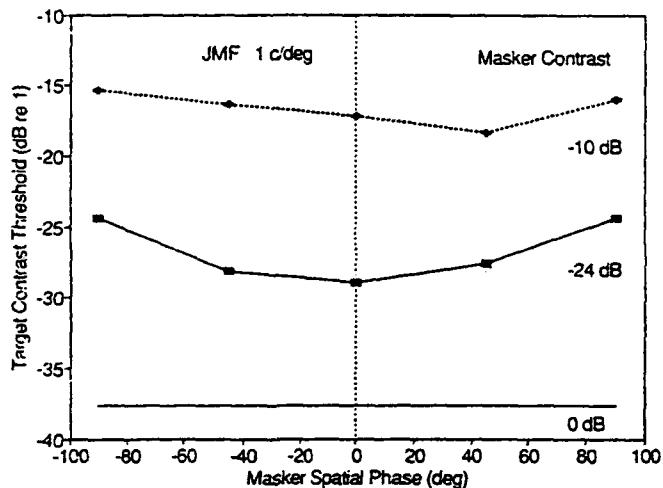


Figure 6. Experiment 2. Threshold versus relative spatial phase of masker at two masker contrasts. Spatial frequency is 1 c/deg.

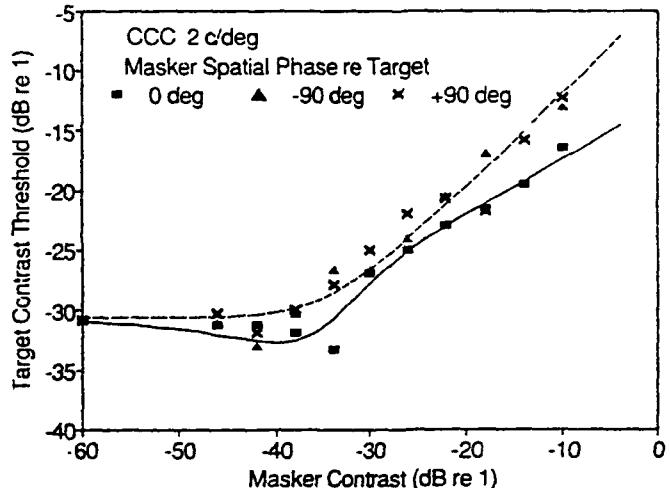


Figure 7. Experiment 3. Tvc functions for maskers with different phase relative to the target.

#### 3.4 Exp. 4: Effect of temporal modulation on Tvc function

Experiment 4 examines simultaneous masking of and by temporally modulated spatial patterns. Here the methods were somewhat different. A spatial forced-choice paradigm was used with the target being centered 1 deg above or below the fixation point. Both target and masker were 2 c/deg and oriented vertically. Maskers had a duration of 667 msec and underwent a sinewave counterphase modulation during that interval. Targets were centered in time with respect to the maskers and underwent Gabor counterphase modulation with a 1/e time constant of 106 msec. Target and masker were spatially and temporally in phase at the center of the stimulus interval. There were two target temporal frequencies and five masker temporal frequencies.

Tvc functions for one observer are shown in Figure 8. On the left are the functions for a target temporal frequency of 10 Hz and on the right, 1 Hz. For a 10 Hz target the functions resemble those for a pulse except that the rising part of the functions are more parallel. Facilitation is present only when target and masker are the same in temporal frequency. For the 1 Hz target there is more deviation from parallelism. Thus multiplicative horizontal displacement fails along the temporal frequency dimension. In each case maximum masking is produced by a masker frequency different from that of the target; 20 Hz is most effective in masking 10 Hz and 10 Hz is most effective in masking 1 Hz. Again the smooth curves correspond to the new theory.

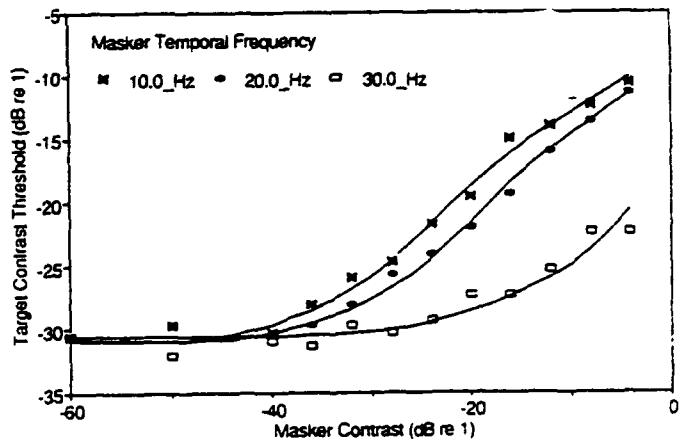
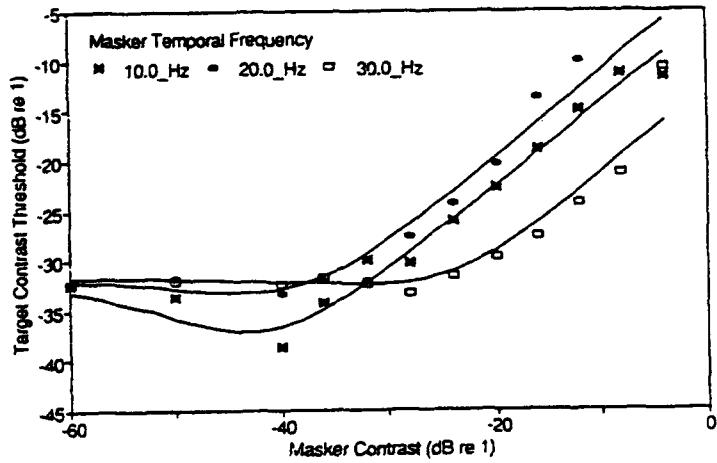
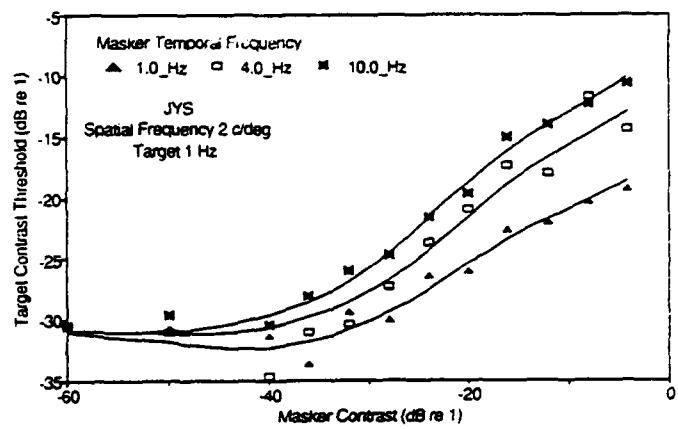
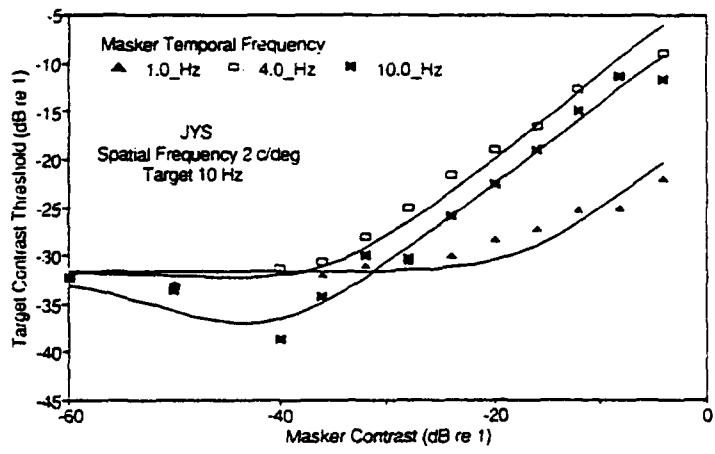


Figure 8. Experiment 4. TvC functions for maskers of different temporal frequencies. Spatial frequency = 2 c/deg.  
a) Target temporal frequency = 10 Hz. b) Target temporal frequency = 1 Hz.

Figure 9 shows the sensitivities estimated from these data using the new theory. On the left are the sensitivity vs temporal frequency functions for the 10 Hz target and on the right for the 1 Hz target. For the 10 Hz target both excitatory and inhibitory sensitivity are bandpass with peaks in the vicinity of 10 Hz. For the 1 Hz target both functions are low pass with inhibitory sensitivity extending to somewhat higher temporal frequencies. In cat cortical cells Bonds<sup>6</sup> found that suppression extended to higher temporal frequencies than excitation.

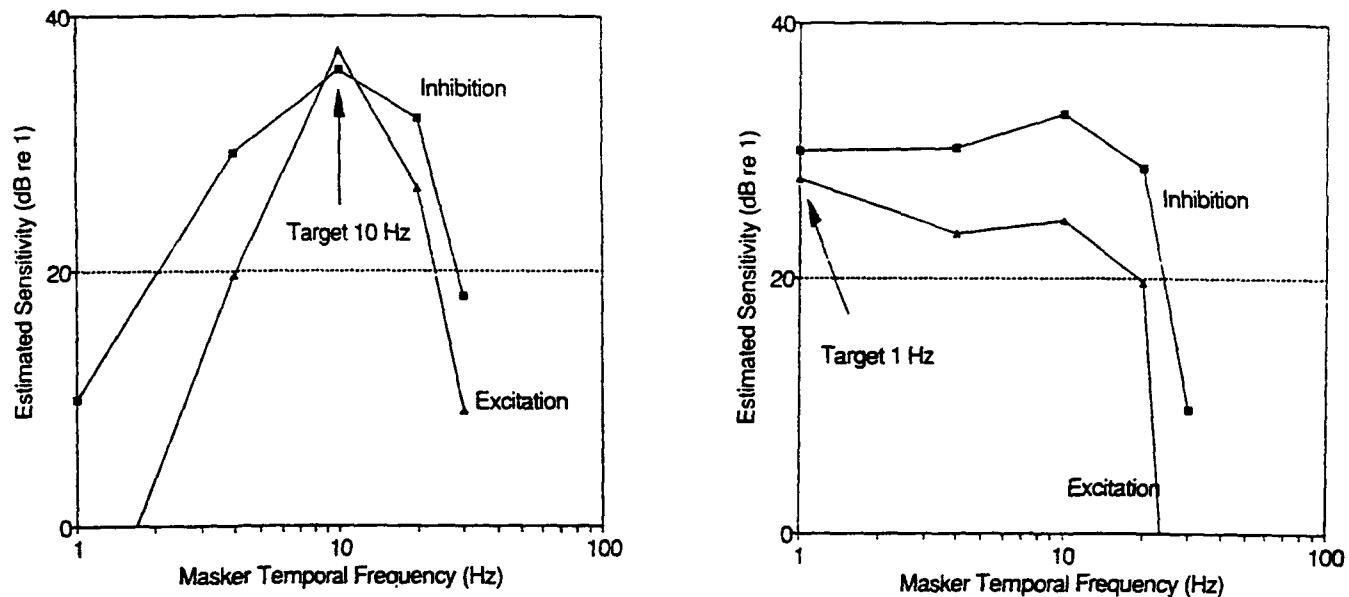


Figure 9. Excitatory and inhibitory sensitivities estimated from the data of experiment 4.

#### 4. DISCUSSION

This same kind of analysis can be applied to masking along other dimensions of the masker such as monocular vs dichoptic presentation<sup>11</sup> and luminance vs chrominance patterns. Although maskers usually excite as well as inhibit, masking of luminance patterns by chrominance patterns appears to be purely inhibitory<sup>12</sup>.

Although the focus here has been on the nature of the human luminance pattern vision mechanisms, the implications concerning the nature of masking are also important. Very early in the history of masking research, two hypotheses were proposed to explain why the masker raises the threshold of the target<sup>13</sup>. One was suppression, which was associated with neural inhibition, and the other was integration, meaning that signals arising from the masker and target get merged into a single signal and this merging somehow makes the target harder to see.

What the new theory asserts is that masking by a similar stimulus involves both suppression and integration, but masking by a remote stimulus involves only suppression. When the masker excites the detecting mechanism, this affects the threshold, but in the new theory the effect of excitation is always to reduce the threshold. Thus, although masker excitation is sometimes involved in threshold determination, it does not

contribute to threshold increase and thus it does not produce simultaneous masking. Simultaneous masking is a consequence of the suppression produced by the divisive inhibitory input to the pattern vision mechanisms. Facilitation, on the other hand, is a consequence of masker excitation.

Recently two new models of the cat simple cell response have been proposed by Heeger<sup>7</sup> and by Albrecht and Geisler<sup>14</sup>. Both of them incorporate a linear excitation process, halfwave rectification, a nonlinear power function transform with an exponent of 2 or greater, and a normalization or contrast gain control process that is influenced by stimuli that do not excite the cell as well as those that do. There are some differences between these models and the model of human pattern vision mechanisms described here, the principle one being that most of the cells saturate within the contrast range studied here, but the human mechanisms do not. Nevertheless, there is a remarkable similarity between the neurobiology and the model that describes human detection performance in the presence of maskers.

### 5. ACKNOWLEDGEMENTS

This paper gives an overview of a research project in which both authors have been involved; each was responsible for some of the experiments. We thank Jerome Tietz for creating the experimental programs, and Kathleen Foley, Julie Shieh and Chien Chung Chen for serving as observers. This research was partially supported by U. S. Public Health Service Grant EY07201 from the National Eye Institute.

### 6. REFERENCES

1. G. E. Legge and J. M. Foley, "Contrast masking in human vision," *Journal of the Optical Society of America*, **70**, 1458-1471, 1980.
2. J. Nachmias and R. V. Sansbury, "Grating contrast: Discrimination may be better than detection," *Vision Research*, **14**, 1039-1042, 1974.
3. D. G. Albrecht and D. B. Hamilton, "Striate cortex of monkey and cat: contrast response function," *Journal of Neurophysiology*, **48**, 217-237, 1982.
4. J. Nachmias, "Masked detection of gratings: the standard model revisited," *Vision Research*, **33**, 1359-1365, 1993.
5. L. A. Olzak and J. P. Thomas, "Seeing spatial patterns," In Boff, K. Kaufman, L. and Thomas, J. P. (eds) *Handbook of Human Perception and Performance*, vol. 1, New York: Wiley, 1986.
6. A. B. Bonds, "Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex," *Visual Neuroscience*, **2**, 41-55, 1989.
7. D. J. Heeger, "Nonlinear model of neural responses in cat visual cortex," In Landy, M. S. and Movshon, J. A. eds., *Computational Models of Visual Processing*, Cambridge, MA: MIT Press, 1989.  
D. J. Heeger, "Normalization of cell responses in cat visual cortex," *Visual Neuroscience*, **9**, 181-197, 1992.
8. A. B. Watson and D. G. Pelli, "QUEST: A Bayesian adaptive psychometric method," *Perception & Psychophysics*, **33**, 113-120, 1983.
9. J. M. Foley, "Human pattern vision mechanisms: Masking experiments require a new theory," submitted.

10. J. M. Foley and G. M. Boynton, "Simultaneous pattern masking: Mechanisms are revealed by threshold versus masker contrast functions and the direct measurement of masking sensitivity," Suppl. to Investigative Ophthalmology and Visual Science, 33, 1256, 1992
11. M. A. Georgeson, "Spatial phase dependence and the role of motion detection in monocular and dichoptic forward masking," Vision Research, 28, 1193-1205, 1988.
12. E. Switkes, A. Bradley, and K. K. DeValois, "Contrast dependence and mechanisms of masking interactions among chromatic and luminance gratings," Journal of the Optical Society of America A, 5, 1149-1162, 1988.
13. B. G. Breitmeyer, Visual Masking: An Integrative Approach. New York: Oxford University Press, 1984
14. D. G. Albrecht and W. S. Geisler, "Motion selectivity and the contrast-response function of simple cells in the visual cortex," Visual Neuroscience, 7, 531-546, 1991.

# Monocular and Binocular Mechanisms of Contrast Gain Control

Izumi Ohzawa and Ralph D. Freeman

University of California, School of Optometry  
Berkeley, California 94720  
E-mail: izumi@pinoko.berkeley.edu

## ABSTRACT

Prolonged stimulation by temporally modulated sinusoidal gratings causes a decrease in the contrast sensitivity and response of neurons in the visual cortex. We have studied the dynamic aspects of this contrast gain control mechanism, and how its temporal properties affect the determination of neural contrast response functions. In addition, we have considered the possibility that a single mechanism is sufficient to explain monocular and binocular properties of contrast gain control.

We find that neural contrast response functions are highly susceptible to the measurement procedure itself so that the data obtained in some studies seriously underestimate the slope of the function and overestimate the threshold. Therefore, careful selection of the experimental data is required for general use and for constructing models of visual cortical function.

Comparisons of monocular and binocular properties of contrast gain control provide insights concerning the neural origin of the mechanism. Monocularly induced gain reductions are transferrable to the other eye, suggesting that gain control originates in part at a site following binocular convergence. However, binocular experiments conducted with interocular contrast mismatches indicate that the gain of the monocular pathways for each eye may be controlled independently. These results suggest that a single gain control mechanism is not sufficient to account for the properties exhibited by cortical neurons.

## 1. INTRODUCTION

Contrast is defined as a ratio of luminances of two separate regions of a visual image. This derivative quantity, however, is the primary parameter that represents the strength of visual stimuli for the visual cortex. This is because the retina, the first stage that senses light and prepares the image for transmission to the cortex, removes much of the information regarding point-by-point luminance distribution in the original image<sup>1,2</sup>. As the retina is highly adaptive to the prevailing absolute luminance levels of the visual scene<sup>2,3,4,5,6</sup>, the visual cortex is highly adaptive to the contrast of stimuli. Therefore, both of these neural structures are dynamic systems whose response characteristics critically depend on their recent history of stimulus exposure. The adaptive mechanism of cortical neurons to the prevailing contrast has been designated contrast gain control<sup>7,8</sup>. In this paper, we illustrate the difficulty of accurately measuring contrast response properties of cortical neurons. We then examine monocular and binocular properties of contrast gain control, and consider organization and possible sites of origin of this mechanism.

## 2. METHODS

Experiments are performed using normal adult cats. Animals are anesthetized during surgery and anesthetized and paralyzed during recording. Details of surgical procedure are presented elsewhere<sup>8,9,10</sup>.

Tungsten-in-glass electrodes are used to record action potentials extracellularly. Spikes are converted into digital pulses and recorded by a computer with 1 msec time resolution. Another computer is used to generate visual stimuli on a pair of displays. For monocular contrast gain control experiments and dichoptic phase shift experiments, high brightness displays ( $250 \text{ cd/m}^2$ , Joyce Electronics) are used. For the cross-orientation inhibition experiments, standard high resolution color monitors are used.

Once a cell is isolated, approximate receptive field location and preferred orientation of the stimuli are determined by a manually controlled bar stimulus. Computer controlled runs are then performed to determine the optimal orientation and spatial frequency quantitatively. Following these runs, specific measurements vary depending on the questions addressed.

### 3. RESULTS

#### 3.1 Monocular properties of contrast gain control

Contrast response functions have received renewed attention recently in quantitative analyses of the visual cortex<sup>11,12,13</sup>. In these analyses, experimental data are used to derive an explicit functional description of a representative contrast response function, thereby allowing estimations of the slope and form of the input-output relationship of these neurons. These estimates are then used to predict tuning characteristics for phase, spatial frequency and degree of direction selectivity<sup>11,12,13</sup>.

Such analyses critically depend on the integrity of the original contrast vs. response data used. Experiments we have conducted show that accurate and reliable data are quite difficult to obtain. This is because a cortical neuron is a component of a dynamic system whose operating point or adaptation level is highly modifiable depending on what stimuli the system has been exposed to in the recent past<sup>7,8,14,15,16,17</sup>. To make matters worse, the standard random interleaving method that is typically used to reduce the effects of intrinsic neural variability causes inaccurate measurements of contrast response functions. This is demonstrated in a series of measurements we have performed for cortical neurons.

Fig. 1 shows results of a series of contrast response function measurements on a complex cell<sup>8</sup>. The dashed curve is the result of randomly interleaved presentations of a wide range of stimulus contrasts ranging from 1.6% to 100%. Stimulus duration for each trial is 4 sec for this run. The solid curves represent contrast response functions obtained by interleaving contrasts of a limited range. By limiting contrasts to a narrow range, the adaptation level remains relatively stable. For example, the left-most curve (open circles) is the result of a run in which contrast was interleaved within the range 1.6% to 6.3% (centered at 3.13% and limited in range to  $\pm 1$  octave). The remaining solid curves were obtained at progressively higher center contrasts, but still within a  $\pm 1$  octave range. A comparison of the maximum slope in each curve shows that curves obtained with a  $\pm 1$  octave contrast range have much steeper slopes than those for the dashed curve. It is also apparent that the curves obtained with limited contrast ranges shift laterally along the log contrast axis. This indicates that the effective dynamic range of the neuron is adjusted to match the range of contrasts in the stimuli. These results indicate that the standard random interleaving technique, as applied to the measurement of contrast response, can seriously underestimate the slope of contrast response functions. Contrast threshold estimates obtained by this procedure are also substantially higher than the cell's actual

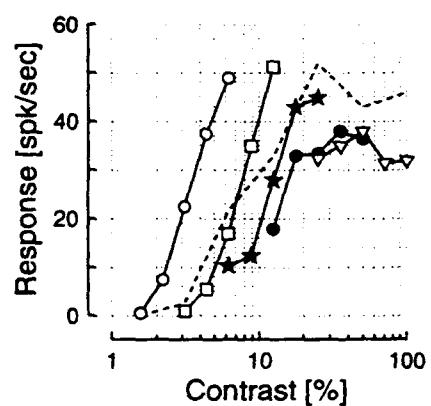


Figure 1

threshold.

Fig. 2 shows another example, for a simple cell, of the difficulty in obtaining accurate contrast response functions. In this test, attempts are made to cause as little adaptation to the cell as possible by employing a brief stimulus presentation (2 sec) preceded by a long period (10 sec) during which no stimulus was present. The result of this test is shown in the left-most curve (open circles). The two curves to the right were obtained for 4-second stimulus presentations with no blank periods between trials. Here again, the two types of measurements yield completely different contrast response functions. With minimal adaptation, the contrast response function on the left (open circles) rises more gradually than the middle curve. The estimated contrast threshold also is much lower when the cell is minimally adapted. These factors clearly affect parameters such as the slope, the exponent of the accelerating nonlinearity, and the threshold that are obtained by curve fitting procedures. Taking the results for Figs. 1 and 2 together, careful selection of the experimental data is required for general use and for constructing models of the visual cortical function.

### 3.2 Binocular properties of contrast gain control

The two questions we address on the mechanisms of contrast gain control for the binocular case are as follows. How is the contrast gain for one eye affected by the level of contrast of stimuli presented through the other eye? Is the site of contrast gain control before or after the binocular convergence of signals from the two eyes? If it is before binocular convergence, the gain control is presynaptic to the cortical neuron under study and should operate monocularly. If the gain control is located after binocular convergence, the site must be postsynaptic and the effects will transfer interocularly. This is certainly the case for interocular transfer of contrast adaption<sup>18</sup>. However the situation may not be exactly the same under dichoptic stimulus conditions. Fig. 3 illustrates schematically a simple model of the gain control mechanism. We incorporate a threshold mechanism for firing that is located postsynaptically. It may be the basis for the gain control that operates following binocular convergence. In this model, the threshold is controlled so that the main section of the contrast response curve will contain the prevailing contrast of the stimuli. A simple prediction based on this model of a common postsynaptic gain control mechanism is that both eyes' gains are affected identically. If either of the two eyes is exposed to a high contrast pattern, the high contrast dictates the adaptation level of the neuron. We have tested this explicitly using dichoptically presented grating stimuli.

The tests consist of a series of dichoptic phase shift experiments we have developed for studying binocular interactions<sup>9</sup>. Fig. 4 shows schematically the stimuli used in a dichoptic phase shift experiment. Here, four dichoptic stimulus conditions are shown as four rows of left-right pairs

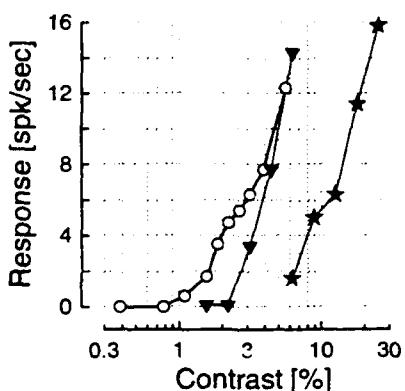


Figure 2

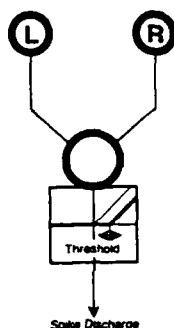


Figure 3

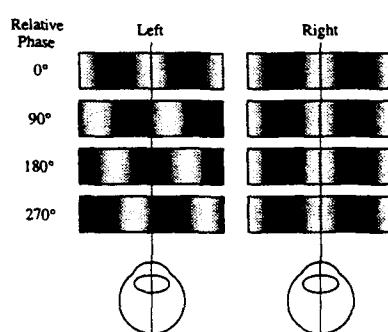


Figure 4

of sinusoidal gratings. Taking the top-most pair as the reference, the grating pairs below have varying degrees of interocular phase shifts in 90° steps. The phase of the stimuli presented to the right eye remains constant while that for the left eye is shifted systematically. In the actual experiments, these grating pairs are moved at a constant velocity in the preferred direction of the neuron while maintaining the relative phase between the left and right gratings. From the perspective of binocular vision, these phase-varying stimuli constitute a complete set of binocular disparity changes if sufficiently small phase steps are used. This is because the grating stimuli are periodic and all cases may be exhausted by varying the relative phase over 360°.

Fig. 5 presents the results of one such dichoptic phase shift experiment. Before this dichoptic phase run, however, the preferred orientation and spatial frequency are determined monoptically for each of the two eyes. Fig. 5A and B show orientation tuning and spatial frequency tuning curves, respectively. The dichoptic phase run is then performed using the optimal orientation and spatial frequency for each eye. Peri-stimulus time histograms (PSTH) are shown in Fig. 5C for all of the conditions measured in a single interleaved run. Relative phase is varied over 360° in 45° steps, and monocular stimulus conditions (depicted as L and R) are also included. From the deep modulation of responses apparent in the PSTH's, this cell is clearly a simple cell. After harmonic analyses of these histograms at the temporal frequency of grating drift, we plot the first harmonic component of the responses in Fig. 5D. To quantify how strongly the input from the two eyes interacts, a cycle of a sinusoid is fit to the binocular response data. From this fit, we are able to determine an index, *depth of modulation*, that quantifies the degree of binocular interaction for this neuron<sup>9</sup>. Fig. 6 shows the definition for this index in which A represents the amplitude of the fitted sinusoid and M is the mean of the binocular responses. According to this definition, the simple cell shown in Fig. 5 had an index of 1.02. In general, the larger the index, the stronger the binocular interaction. An index of nearly zero indicates that the binocular interaction curve is almost flat. This means that stimuli presented to one eye had no influence on the responses elicited by the other eye. Therefore, the depth of modulation quantifies the influence one eye has on a neuron using the excitation through the other eye as the reference level.

Using these analysis methods, we now address the following question: Is the contrast gain of a neuron dictated by the eye that receives higher contrast when left and right eyes are stimulated by gratings of unequal contrast? To answer this question, we perform a series of dichoptic phase shift experiments with varying

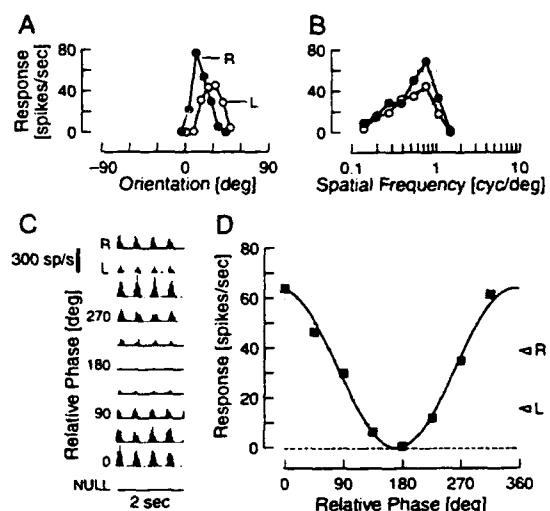


Figure 5

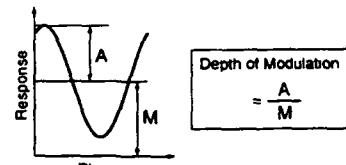


Figure 6

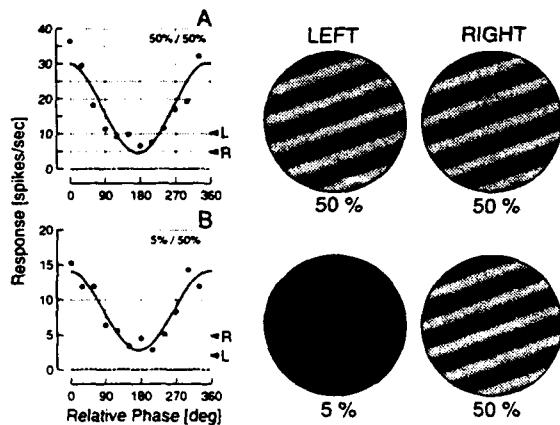


Figure 7

combinations of contrast mismatch between the two eyes. Fig. 7 illustrates the stimuli and presents the results for two extreme conditions for a simple cell. In Fig. 7A, a control condition is shown with equal contrast gratings (50% each) presented to the two eyes. The result is very similar to that shown in Fig. 5D, except that in this run 12 relative phase values are used in steps of 30°. Fig. 7B presents the results for the condition in which the left eye was stimulated with a grating of 5% contrast which is one log-unit lower than the contrast for the right eye, maintained at 50%. The difference in contrast, as illustrated by the grating patches shown on the right, is striking on inspection. To our surprise, the dichoptic phase run reveals a binocular interaction that is nearly identical to that of Fig. 7A with equal contrasts for the two eyes. The degree of binocular interaction as quantified by the depth of modulation index remains unchanged even with a one log-unit contrast mismatch. This is surprising because our initial prediction, based on the postsynaptic model of contrast gain control, that is located after binocular convergence, is a dramatic reduction of binocular interaction for dichoptic stimulation with grossly mismatched contrasts. With a ten-fold decrease in left eye stimulus contrast, we had expected a similar decrease in the depth of modulation. The peak discharge rate and the peak-to-trough response difference are indeed lower for the mismatched contrast condition. However, this difference is still much smaller than the change in the left eye contrast. For the matched contrast condition, the peak-to-trough response difference is 25 spikes/sec, while the mismatched contrast condition produces a difference of 11 spikes/sec. This accounts for a difference of only 2.3 times, compared with the 10-fold difference in left eye contrast. Results from all the dichoptic measurements for this cell are summarized in Fig. 8. In this figure, both depth of modulation indices and monocular response levels are plotted against the contrast of the grating presented to the left eye. The contrast for the right eye was held constant at 50%, and response level to this stimulus is shown by the triangle. It is clear that the depth of modulation (filled circles) remains constant at nearly 0.7 for all the conditions. Responses for stimulation of the left eye alone, however, are critically affected by the stimulus contrast as expected (open circles).

Results for a similar series of dichoptic phase shift experiments performed on a complex cell, are shown in Fig. 9, in a format similar to that of Fig. 7. Again, despite nearly a 1 log-unit difference in interocular contrast, the degree of binocular interaction is maintained at a constant level. A summary of all measurements performed on this cell is shown in Fig. 10. The depth of modulation remains at about 1.0 for a wide range of contrast mismatches (filled circles). Monocular responses to stimulation of the left eye again show expected contrast dependence (open circles).

The two cells presented above are not unique in their behavior. We have conducted similar sets of dichoptic runs on a total of 21 cells, and a large majority of them show constancy of binocular interaction under interocular contrast mismatch conditions. Fig. 11 shows the results for all the

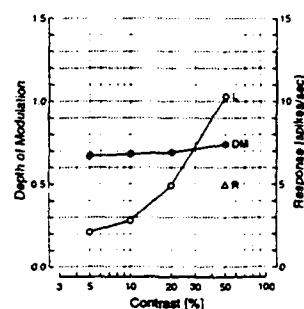


Figure 8

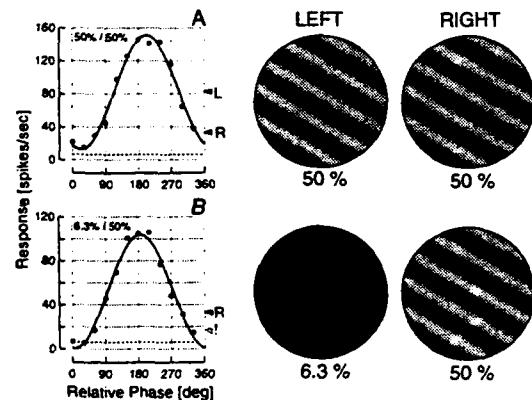


Figure 9

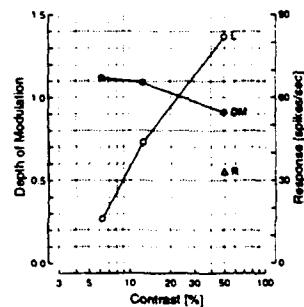


Figure 10

cells we have studied. In Fig. 11A, the depth of modulation index for each cell is plotted against the left eye contrast. The contrast for the right eye was maintained at 50%. It is clear that more than half of these cells maintain relatively constant binocular interaction for a wide range of interocular contrast mismatch even, in some cases, with a 20-fold difference in contrasts (2.5% and 50% for the left and right eyes, respectively). The mean and the spread of the data in Fig. 11A are plotted in 11B. The length of the error bars (from the mean to each end) represents 1 standard deviation. The results shown above indicate that the eye that is presented only a very low contrast is still able to influence the responses of cells strongly. Often a stimulus presented to one eye that is 10 to 20 times weaker than that presented to the other can completely suppress the response from the cell. These results indicate that the postsynaptic mechanism of contrast gain control alone, in the form shown in Fig. 3, is not sufficient to account for the behavior of cells studied under mismatched interocular contrasts. To reconcile the results, it appears that another mechanism is required at a presynaptic site before binocular convergence, in addition to the postsynaptic mechanism which is binocular. This and the implications of the constancy of binocular interaction are discussed below (section 4).

In a separate line of study, we have examined another type of effects, *cross-orientation inhibition*<sup>10,19,20</sup>, which appears to be a form of gain control, that operates monoptically, but not dichoptically. In this effect, responses to an excitatory stimulus, usually of optimal orientation, are suppressed by a superimposed stimulus of another orientation. The panels at the top of Fig. 12 illustrate a typical stimulus configuration. The receptive fields of a cortical neuron are represented by dashed squares. The preferred orientation of the cell is indicated by dashed lines going through the middle of the receptive field. In a typical experiment, a grating of preferred orientation is presented to the receptive field. Then, the response strength to this stimulus is compared with that elicited by a stimulus that is composed of two superimposed gratings, one at the optimal orientation and the other orthogonal to it. The top-left panel of Fig. 12 shows this latter stimulus. For nearly all cells, the response to the latter composite stimulus is weaker than that to the excitatory stimulus alone.

One of the questions we addressed for the cross orientation inhibition is whether there is an effect when the inhibitory (usually orthogonal) grating is presented to the other eye instead of to the same eye that receives the excitatory grating. This dichoptic stimulus configuration is depicted at the bottom of Fig. 12. The question is important because the answer to it can reveal the site of the phenomenon, whether it is presynaptic or postsynaptic to the cortical neuron under study. If the effect is present dichoptically, it indicates that the source of the inhibitory signal is binocular. If, on the other hand, the phenomenon is monocular, it must mean that both the source of the inhibitory signal and the site of action of the inhibition must be presynaptic. Fig. 13 presents the results of this experiment. Three cases are shown: (1) dichoptic condition where the excitatory grating is presented to one eye and the inhibitory (orthogonal) grating is presented to the other (filled triangles), (2) both the excitatory and orthogonal gratings are presented to the

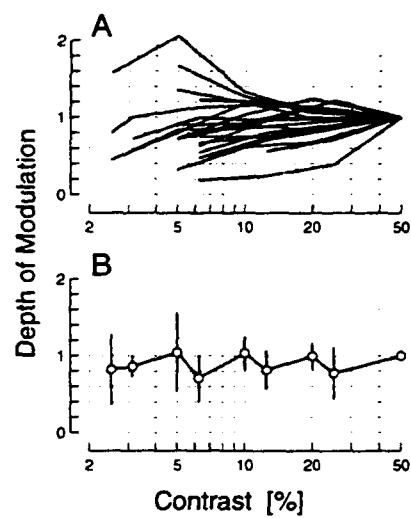


Figure 11

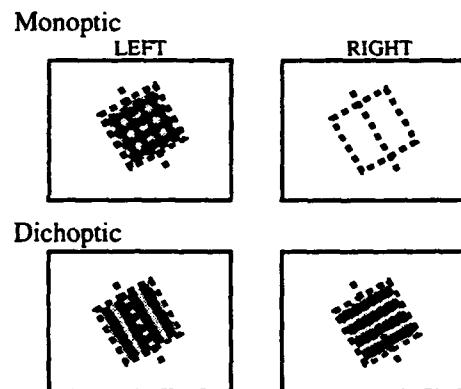


Figure 12

left eye (open circles), (3) both gratings are presented to the right eye (open squares). In these measurements, the excitatory grating is presented at optimal parameters, while the spatial frequency of the orthogonal grating is varied within a range from 0.14 - 2.6 cycles/deg. The response levels to the excitatory grating alone, obtained from a trial interleaved with those for two gratings, are plotted along the right edge of the figure. These monocular responses serve as a control against which the suppression is measured. It is clear that monoptic presentations of both the excitatory and inhibitory gratings have effects which depend on the spatial frequency of the orthogonal grating. The inhibition is strongest at about 0.3 to 0.6 cycles/deg, and essentially disappears at frequencies above 2 cycles/deg. For the dichoptic case (filled triangles), no suppressive effect is observed. The results indicate, therefore, that cross-orientation inhibition acts only monoptically when the inhibitory stimulus is presented to the same eye as that receiving the excitatory optimal stimulus. Therefore, these results indicate that the inhibitory signal must originate from monocular sources, and must act in the visual pathway that is still monocular, i.e., presynaptic to the neuron from which we are recording.

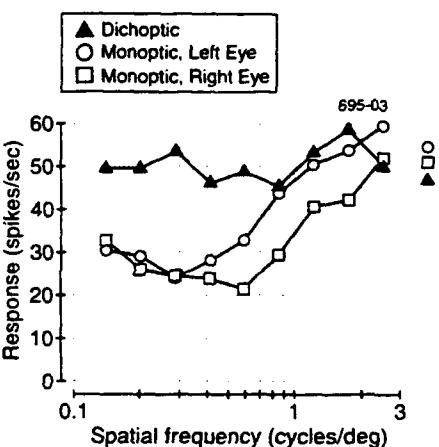


Figure 13

#### 4. DISCUSSION

In this paper, we have examined monocular and binocular properties of dynamic contrast gain control. In the first part, we have shown that a random interleaving of stimuli may seriously interfere with the intended goal of accurately measuring a contrast response function. In most physiological measurements, randomly interleaved tests of multiple stimulus conditions in a single run reduce the effects of inherent variability in neural responses. However, when the stimulus parameter itself drives the operating point (adaptation level) of the neuron, appropriate measures must be taken to guarantee that the stimulus vs. response curve represents a meaningful relationship. We have demonstrated that, when care is taken to minimize the fluctuation of the adaptation level by using a limited range of contrast levels, the slope of the contrast response functions is much more steep and contrast threshold is substantially lower than when a wide range of contrast values are interleaved in a single run. Although this has been known for quite some time<sup>7,8</sup>, there have been a number of studies that still employ the stimulus interleaving technique for the wrong stimulus variable, i.e., contrast. Since not all the data available in the literature represent accurate contrast vs. response relationships, one must be selective in choosing previous results for use in quantitative analyses of contrast response functions.

In the second part, we have examined monocular and binocular aspects of contrast gain control. Our main result here is that there appear to be at least two mechanisms of contrast gain control. A schematic notion of a plausible organization of these mechanisms is illustrated schematically in Fig. 14. In this scheme, a common form of contrast gain control acts at a site after binocular convergence. This

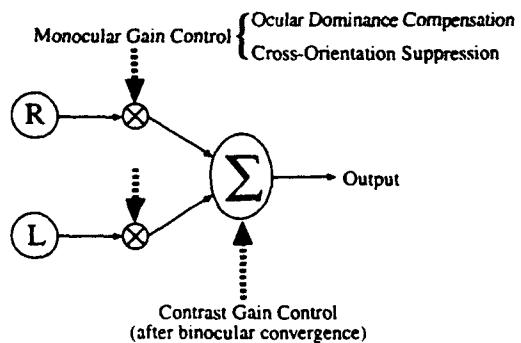


Figure 14

mechanism can account for contrast adaptation that exhibits interocular transfer<sup>18</sup>. The other type of contrast gain control operates monocularly. In our scheme, two cases are illustrated in which monocular forms of contrast gain control operate: (1) maintained binocular interaction even under large interocular contrast mismatches, (2) cross-orientation inhibition. It is of interest to know if these two monocular forms of contrast gain control are the result of a common mechanism. Unfortunately, we are unable currently to determine if this is the case.

One interesting speculation is possible for the role of maintained binocular interaction under contrast mismatch. This could serve as a mechanism that compensates for natural variations of ocular dominance<sup>21</sup>. The rational is as follows. The phenomenon provides an extremely stable system that maintains the effectiveness of input from both eyes despite large variations in the relative strengths of signals from the two eyes. In our experiments, these variations in the relative signal strength between the eyes are created artificially by manipulating the contrast of stimuli for the two eyes, and the experiments are possible only for binocularly responsive cells. In the visual cortex, there are large intrinsic variations of ocular dominance from one cell to another. These ocular dominance variations create exactly the same effect as that produced by the manipulation of stimulus contrasts, because both create mismatched input strengths in the two monocular pathways converging onto a postsynaptic neuron. Therefore, if the monocular presynaptic mechanism of gain control acts to amplify weak input to maintain its influence on the cell, the same mechanism should also function as one that compensates for ocular dominance differences. A consequence of such a mechanism is that nearly all neurons will be able to maintain effective connections from the two eyes regardless of their ocular dominance. In this sense, ocular dominance variation may be an insignificant consequence of developmental process that the brain is able to nullify under binocular operating conditions.

### 5. ACKNOWLEDGMENTS

This work was supported by research and CORE grants from the National Eye Institute (EY01175 and EY03176).

### 6. REFERENCES

1. Kuffler, S. Discharge patterns and functional organization of mammalian retina. *J. Neurophysiol.* **16**: 37-68, (1953).
2. Werblin, F.S. Control of retinal sensitivity. II. Lateral interactions at the outer plexiform layer. *J. Gen. Physiol.* **63**: 62-87, (1974).
3. Laughlin, S.B. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.* **36**: 910-912, (1981).
4. Creutzfeldt, O.D. Transfer function of the retina, *Electroencephalogr. Clin. Neurophysiol. Suppl.* **31**: 159-169, (1972).
5. Norman, R.A. and Werblin, F.S. Control of retinal sensitivity. I. Light and dark adaptation of vertebrate rods and cones. *J. Gen. Physiol.* **63**: 37-61, (1974).
6. Werblin, F.S. and Copenhagen, D.R. Control of retinal sensitivity. III. Lateral interactions at the inner plexiform layer. *J. Gen. Physiol.* **63**: 88-110, (1974).
7. Ohzawa, I., Sclar, G., and Freeman, R.D. Contrast gain control in the cat visual cortex. *Nature* **298**: 266-268, (1982).

8. Ohzawa, I., Sclar, G., and Freeman, R.D. Contrast gain control in the cat's visual system. *J. Neurophysiol.* **54**: 651-665, (1985).
9. Ohzawa, I. and Freeman, R.D. The binocular organization of simple cells in the cat's visual cortex. *J. Neurophysiol.* **56**: 221-242, (1986).
10. DeAngelis, G.C., Robson, J.G., Ohzawa, I., and Freeman, R.D. Organization of suppression in receptive fields of neurons in cat visual cortex. *J. Neurophysiol.* **68**: 144-163, (1992).
11. Heeger, D.J. Modelling simple cell direction selectivity with normalized, half-squared, linear operators *Invest. Ophthalmol. Visual Sci. Suppl.* **33**: 953, (1992).
12. Albrecht, D.G., and Geisler, W.S. Motion selectivity and the contrast-response function of simple cells in the visual cortex. *Visual Neurosci.* **7**:531-546, (1991).
13. DeAngelis, G.C., Ohzawa, I., and Freeman, R.D. Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. linearity of temporal and spatial summation. *J. Neurophysiol.* **69**: 1118-1135, (1993).
14. Vautin, R.G. and Berkley, M.A. Responses of single cells in cat visual cortex to prolonged stimulus movement: neural correlates of visual aftereffects. *J. Neurophysiol.* **40**: 1051-1065, (1977).
15. Albrecht, D.G., and Hamilton, D.B. Striate cortex of monkey and cat: contrast response function. *J. Neurophysiol.* **48**:217-237, (1982).
16. Albrecht, D.G., Farrar, S.B., and Hamilton, D.B. Spatial contrast adaptation characteristics of neurones recorded in the cat's visual cortex. *J. Physiol. (London)* **347**: 713-739, (1984).
17. Bonds, A.B. Temporal dynamics of contrast gain in single cells of the cat striate cortex. *Visual Neurosci.* **6**: 239-255, (1991).
18. Sclar, G., Ohzawa, I. and Freeman, R.D. Contrast gain control in the kitten's visual system. *J. Neurophysiol.* **54**:668-675, (1985).
19. Morrone, M.C., Burr, D.C., and Maffei, L. Functional implications of cross-orientation inhibition of cortical visual cells. *Proc. R. Soc. Lond. B Biol. Sci.* **216**: 335-354, (1982).
20. Bonds, A.B. Role of inhibition in the specification of orientation selectivity of cells in the cat striate cortex. *Visual Neurosci.* **2**: 41-55, (1989).
21. Hubel, D.H. and Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol. (London)* **160**: 106-154, (1962).

## **COMPUTATIONAL RECONSTRUCTION OF THE MECHANISMS OF HUMAN STEREOPSIS.**

Christopher W. Tyler, Lauren Barghout and Leonid L. Kontsevich.

Smith-Kettlewell Eye Research Institute  
San Francisco CA 94115 USA.

### **Abstract**

The properties of human stereoscopic mechanisms may be derived from dichoptic interaction and masking effects on stereoscopic detection thresholds in any relevant stimulus domain (spatial frequency, temporal frequency, disparity, orientation, etc.). The present study focuses on the spatial properties of mechanisms underlying stereoscopic depth detection. The computational approach is based on the full exploration of plausible model structures to characterize their idiosyncrasies, which often allows exclusion of proposed mechanisms by comparison with data obtained under conditions in which the idiosyncrasies should be expressed.

For example, we conducted a detailed analysis of threshold elevation functions (TEFs) under plausible channel shapes, combination rules and masking behavior derived from previous studies (e.g., Blakemore and Campbell, 1969; Quick, 1974; Stromeier and Klein, 1974; Legge, 1981; Wilson, McFarlane and Phillips, 1983). The analysis reveals that TEFs may be much narrower than and differ in shape from the underlying mechanisms. For example, only two discrete channels are required to produce TEFs peaking close to each fixed test frequency, with no relation to channel peaks.

We apply this analysis to the stereospatial masking functions collected by Yang and Blake (1991) to determine the likely channel structure underlying the empirical masking performance. The analysis generally supports the two mechanism model that they propose but shows that the assumptions underlying their estimates of the unmasked sensitivity function are incorrect. The analysis excludes stereospatial channels tuned below 2.5 c/deg, a region in which Schor, Wood and Ogawa (1984) obtained evidence for many narrowly tuned channels by measuring disparity thresholds for targets with different peak tunings in the two eyes. Our computational model for the latter data is consistent with the lowest tuned channel being at 2.5 c/deg, this channel being narrowly tuned to dichoptic contrast differences, as described by Legge and Gu (1989) and Halpern and Blake (1988). Thus, all such stereo tuning data can be explained in a model in which all stereoscopic channels are tuned above 2.5 c/deg.

### **INTRODUCTION**

Threshold elevation functions (TEFs) have been measured in many domains of psychophysics by a variety of both adaptation and masking paradigms. In an adaptation paradigm, the observer first measures detection thresholds across the range of a stimulus continuum (such as orientation, spatial frequency, chromatic wavelength, etc.). The observer then adapts to the prolonged presentation of a stimulus at one point on the continuum, then tests detection threshold at the same or a different point on the continuum. An example of this procedure for the domain of the spatial frequency of sinusoidal luminance gratings is shown in Fig. 1. The ratio of the adapted threshold to the unadapted threshold (or some modification of that ratio) provides the measure of threshold elevation by the adaptation process.

## IMPLICATIONS OF DISCRETE CHANNEL MODELS

Historically the first channel modeling in vision was with discrete channel models, as exemplified by the threshold elevation paradigm originally developed in color vision by Stiles (1939). Discrete channel analysis of TEFs in spatial vision goes back to Wilson and Bergen (1979) and has been applied in an initial to TEFs for the spatial structure underlying stereopsis (Yang and Blake, 1991). It has also been used for TEFs in a variety of other stimulus domains, such as temporal frequency (Anderson and Burr, 1985; Hess and Snowden, 1992), motion (Anderson and Burr, ???) and stereomotion (Beverley and Regan, 1973).

The simplest discrete channel case to analyze is that of two channels, for which we assume the power DoG form given by

$$\text{DoG}^p = \left( e^{-\left(\frac{x}{\sigma_1}\right)^2} - e^{-\left(\frac{x}{\sigma_2}\right)^2} \right)^p \quad (1)$$

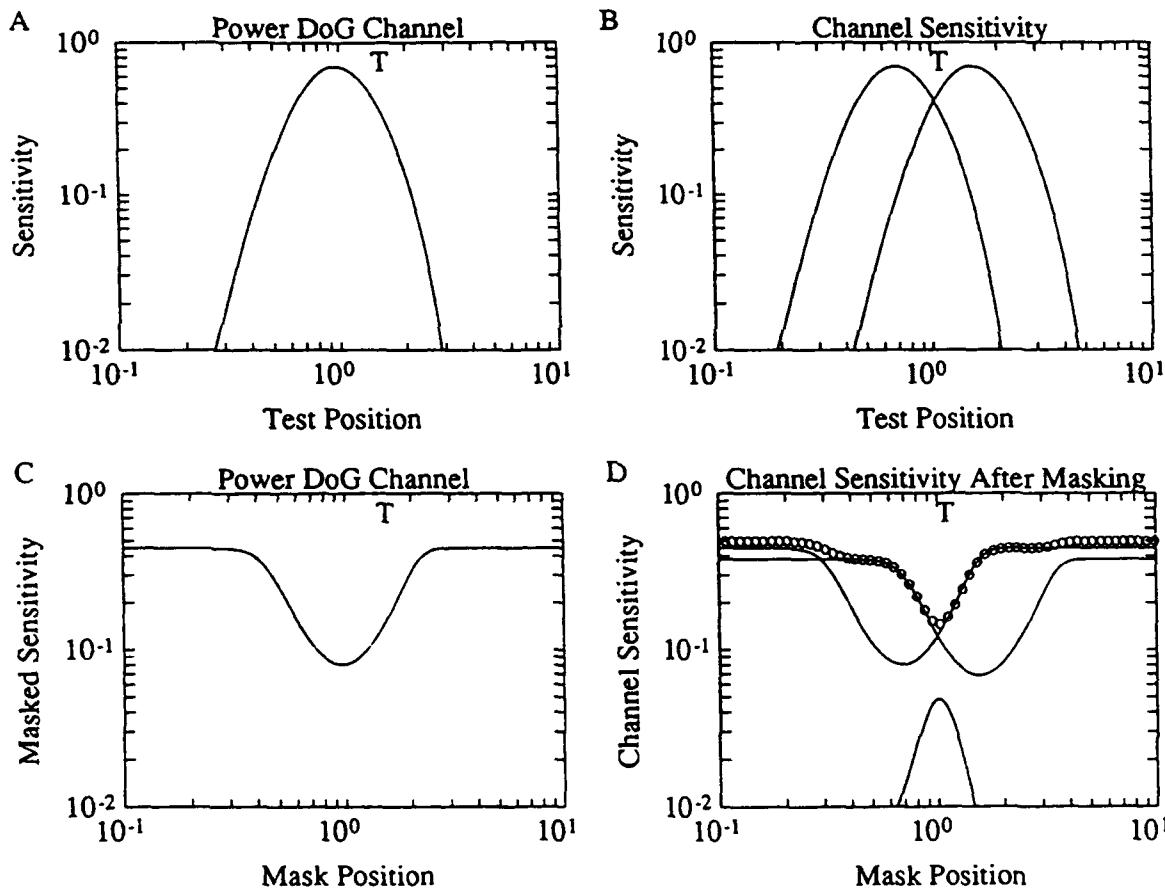


Fig. 1. Fixed-test masking effects for two-channel case.

- Single power DoG channel profile.
- Overlapping pair of power DoG channel profiles. Test stimulus is just to left of intersection.
- Masked sensitivity for single channel in A.
- Upper full curves - masked sensitivities for pair of channels in B. Chain curve - combined sensitivity with a probability summation exponent of 4. Lower curve - TEF that peaks at the test position (+) does not reflect the peak positions or bandwidths of either channel.

The masking function was assumed to be

$$M_\omega = \left\{ \left[ 1 + r_\omega (a_\omega z_m)^2 \right]^\mu \right\}^{-1/2} \quad (2)$$

with the exponent  $\mu$  set to 1. The masking sensitivity function for a single channel and a fixed test stimulus, as mask position on the relevant stimulus dimension is varied, is shown in Fig. 1A&C. Fig. 1B shows a pair of such channels separated by one bandwidth. Notice that this individual masking sensitivity function does not peak at the test position but falls at the peak of the channel sensitivity function. When the two adjacent channels of Fig. 1B are subjected to the same fixed-test masking paradigm, they combine in the fashion depicted in Fig. 1D. For this test stimulus position, the left channel is the most sensitive of the two when the mask is out of range of either channel (i.e., at far left or right of the figure). As the mask is swept toward the position of the test stimulus, the left channel becomes progressively masked so that the residual sensitivity of the right channel becomes dominant. As the mask sweeps past the test position, the sensitivities of the two channels again cross over so that the left channel becomes the most sensitive again. The left channel retains this dominance for the rest of the range of mask positions.

The channel responses  $r_\omega$  in the stimulus domain  $\Omega$  are combined through the standard probability summation equation, or norm, to produce the overall predicted response  $r$

$$\|r\|_\beta = \left( \sum_{\omega \in \Omega} |r_\omega|^\beta \right)^{1/\beta} \quad (3)$$

to produce the overall masked sensitivity function (chain curve in Fig. 1B). Note that this function has its minimum at the test position rather than at the peak of either of the channels.

The same property is exhibited by the TEF derived from this overall sensitivity function by inversion from the internal sensitivity function to the measured threshold elevation prediction. This TEF (Fig. 1D) is also much narrower around its peak than are the channel profiles. Both properties arise because the masked overall sensitivity function is derived from the *intersection* between the two channel profiles (rather than their union, as is the case for the unmasked overall sensitivity function). This figure makes it clear that there is no direct relationship between the TEF and the sensitivity function of either of the underlying channels, either in form, bandwidth or position. Note that, as in the case of continuous channel summation, the precise summation rule will have only minor effects on the shape of the overall function. For example, a change from a summation exponent of 4 to 2 would only serve to raise the overall masked sensitivity function up by a small amount, but would not materially alter its shape.

A further feature of the TEF for two discrete channels is that its shape is subject to idiosyncratic inflections and shoulders. Although this is evident to some extent in Fig. 1 B & D, it is illustrated more clearly in Fig. 2. The only change between the two figures is that the test position has been shifted by a small amount to the right. This shift has the effect of making the right channel the more sensitive of the two, which introduces a crossover between

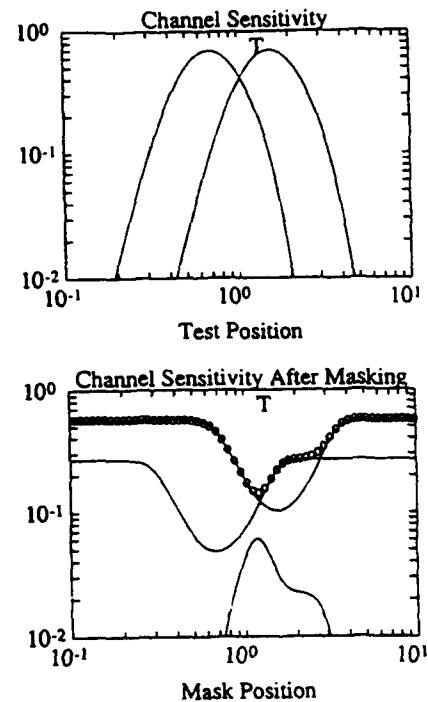


Fig. 2. Effect of shifting the test position for channel pair in Fig. 1

the two functions at a position to the right of the test position. The consequence is a striking change in the form of the TEF to have a pronounced shoulder on the right, where the function previously fell smoothly down from the peak. Thus, TEF functions with inflections and shoulders are to be expected in processing systems containing only a few channels spaced widely apart relative to their bandwidths. These irregularities are result of the discrete structure of the underlying channels, and are not related to any nonlinearity of the masking amplitude function, which was assumed to be linear for Figs. 1&2. (Corresponding irregularities have been noted by Georgeson and Harris, 1984, in their slope function analysis.). Such deviations from smooth TEFs are evident in published data (notably those of Blakemore and Campbell, 1969), but it is beyond the scope of the present exposition to attempt to derive the channels that might be responsible for them.

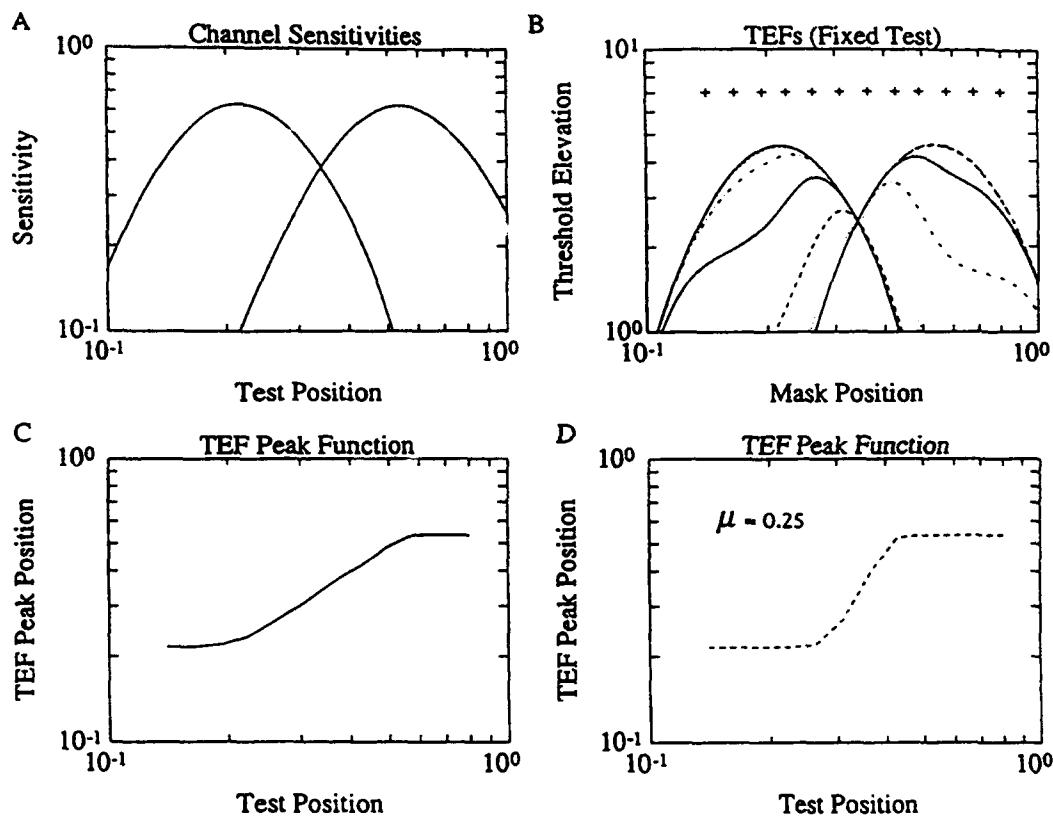


Fig. 3. Continuum of TEF peaks as a function of position of test stimulus in fixed-test paradigm.  
A. Pair of channels with moderate overlap.

B. Array of TEFs (full curves) for a set of test positions (Os) for the two channels in A.

C. Positions of TEF peaks in B as a function of the position of test stimulus.

D. Steeper slope of the TEF peak function with  $\mu = 0.25$  in masking function.

#### Continuum of Fixed-Test TEFs from a Two-Channel Model

The implication of the two-channel model results of Figs. 1&2 is that the measured TEF will peak at the position of the fixed test stimulus, as long as the test falls within the range defined by the channel peaks. This result is illustrated in Fig. 3, which shows the relation of peak TEF to test position for the fixed-test paradigm. Note that the curve in the central range is quite smooth and adheres to the line of strict proportionality over a range of about 2 octaves, the bandwidth of the underlying channel profiles. This function therefore emphasizes that there may be a perfectly continuous relationship between the peak TEF and test position even for a model as discrete as one consisting of just two channels. However, the fact

that the ends of the functions run horizontal indicates the points where this relationship fails and the peak TEF asymptotes to the peak positions of the channels themselves.

The full proportionality of TEF peak to test position is disturbed if the assumption of linearity of the masking amplitude function is violated. It is only in the case of that linearity that the masked sensitivity functions of Figs. 1&2 will cross over exactly at the test position. (In fact, if linearity pertains, that crossover will remain stable even when the underlying channels have different sensitivities. It is the

linearity of the masking amplitude function that has the effect of exactly compensating any difference in the channel sensitivities at the test position with the reciprocal difference in masking effect, so that the masked sensitivities are precisely equated at the test position.) A departure from linearity, such as the power-law saturation implied by an exponent less than 1 in eq. 2, will mean a failure in the reciprocity between test and mask sensitivity differences at the test position, with a consequent shift of the TEF peak toward the more sensitive channel.

TEFs measured for discrete channel systems are largely unaffected by the summation rule by which the channels are summed together to produce the overall output. As shown by Fig. 4, a TEF measured with intrusion from neighboring channel peaks, such as that in Fig. 2, will maintain the stepped peak structure all the way from  $\beta = 1$  (linear summation of the channel profiles) to  $\beta = \infty$ .

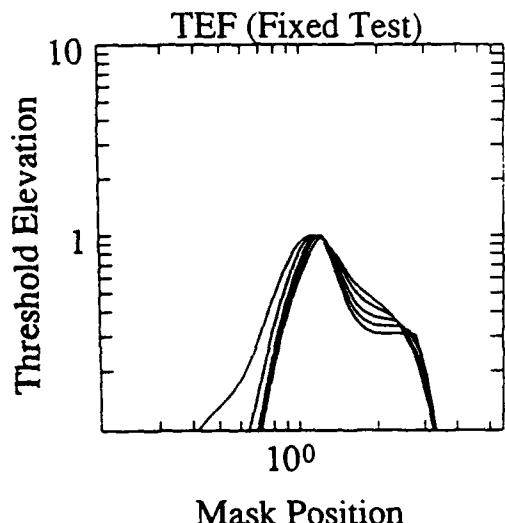


Fig. 4. Effect of summation rule of eq. 3 on TEF shape for discrete channel model of Fig. 1. Curves from top to bottom correspond to TEFs with  $\beta = 1, 2, 4, 8$ , and  $\infty$ .

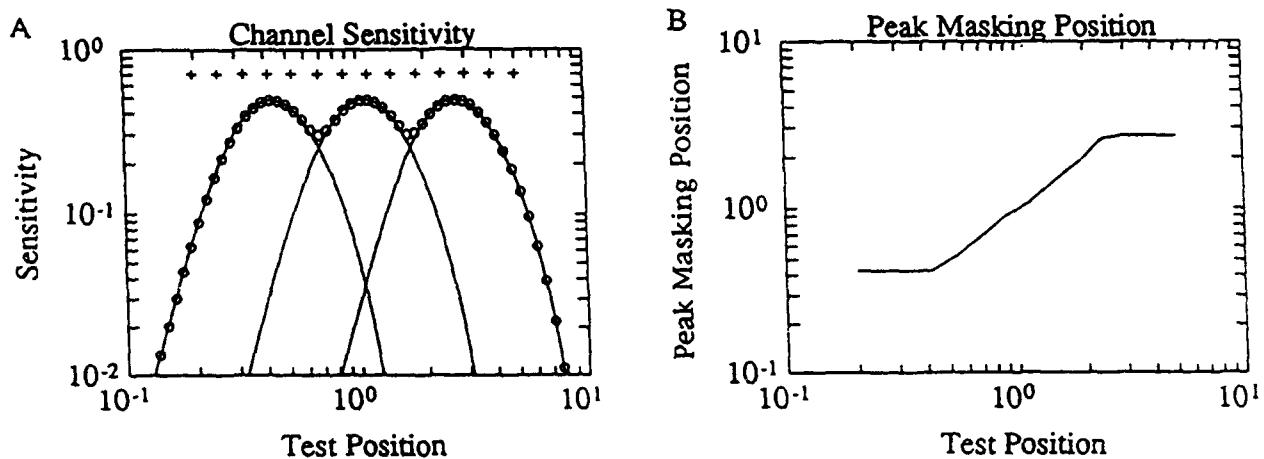


Fig. 5. Visibility of channel discreteness in unmasked sensitivity function.

- A. Triplet of overlapping channels (full lines). With  $\beta$  of 4, unmasked sensitivity function (chain curve) shows pronounced dips where channels intersect.
- B. Peak masking function for channels in A. Note that function is smooth and shows no hint of discrete channel structure.

### Role of Unmasked Overall Sensitivity Function in Revealing Channel Discreteness

The previous section emphasized that a continuum of TEF peaks, which has generally been considered a hallmark of a continuous underlying channel distribution, is instead what should be expected for a discrete channel array of as few as two channels. How then could one distinguish between continuous and discrete channel distributions on the basis of TEF measurements? One answer is provided by the unmasked overall sensitivity curve, which now may be seen to be significantly more sensitive to the discreteness of the channel distribution than the TEF peak function. Consider a triplet of channel profiles separated by their full bandwidth (Fig. 5A). The unmasked overall sensitivity curve has two clear valleys because the probability summation process tracks closely the profile of the most sensitive channel at any point. By extension from Fig. 4, on the other hand, the TEF peak function is again perfectly smooth for the case of  $\mu = 1$  (full line in Fig. 5B). A discrete channel structure therefore should not be expected to be revealed in irregularities in the peak masking function for the fixed-test paradigm, as has often been implied (e.g., Wilson et al., 1983; Lehky, 1985; Swanson and Wilson, 1985; Hess and Snowden, 1992). Careful measurement of detection sensitivity in the absence of masking, on the other hand, may provide a sensitive measure of the presence of discrete channels underlying the visual response across a stimulus domain.

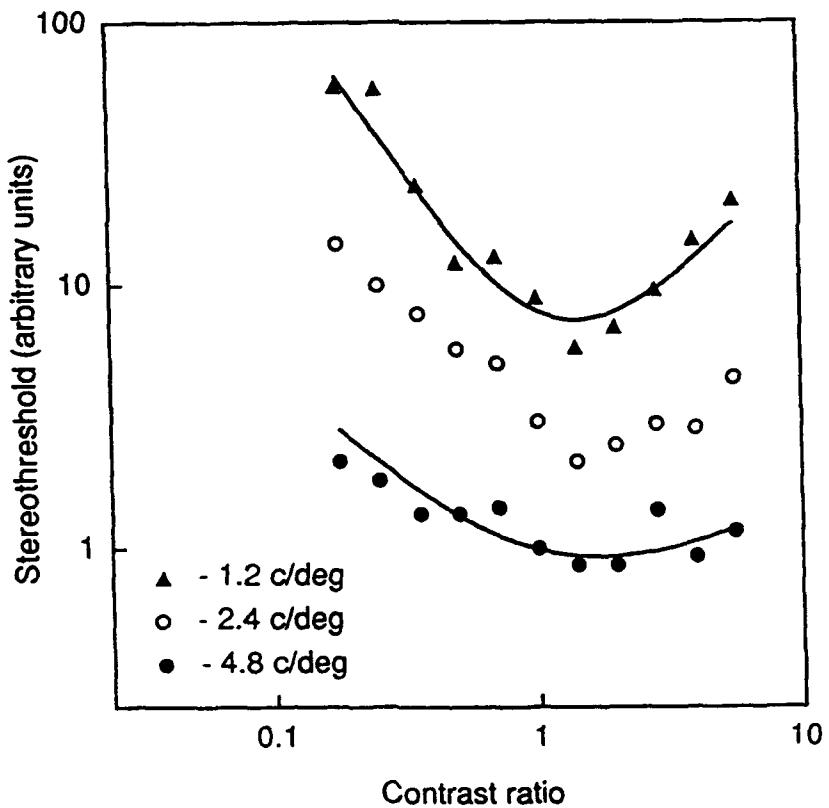


Fig. 6. Dichoptic contrast tuning functions for D10 stimuli with central frequencies 1.2, 2.4 and 4.8 cy/deg derived from Halpern and Blake (1988). The two parts were measured separately for the contrast ratios above and below a value of one in vertical direction. Data points are represented by different symbols; solid lines are prediction of the model.

## THE SPATIAL STRUCTURE OF STEREOPSIS

The core information about the structure of spatial channels underlying stereopsis comes from published data on disparity detection thresholds for a wide range of conditions (Halpern and Blake, 1988; Legge and Gu, 1989; Schor and Heckmann, 1989; Schor and Wood, 1983; Schor, Wood and Ogawa, 1984; Yang and Blake, 1991). This type of detection dramatically changes behavior at the spatial frequency of 2.5 cy/deg; above this frequency all the thresholds remain constant while below it they grow at a uniform rate. Many other thresholds, such as upper disparity limits, threshold amplitudes for stereo and monocular motion show similar behavior. These data lead to the postulate that the fall-off occurs because there are no stereo channels peaking below 2.5 cy/deg, so that the stimuli in the whole range below 2.5 cy/deg are processed by the single channel tuned to 2.5 cy/deg. Consequently, disparity detection thresholds at frequencies below 2.5 cy/deg are controlled by the single parameter of effective contrast at the 2.5 cy/deg channel, whose output depends jointly on the contrast and spatial frequency of the stimuli. We develop this idea to explain the relations between spatial and contrast tuning functions for disparity thresholds. To validate our conclusions we describe an experiment with difference of Gaussian stimuli over a range of interocular widths and contrasts. For a dichoptic width ratio of 2:1, the dichoptic contrast ratio required to minimize disparity detection thresholds was 1:4, just as predicted by the model.

There are four key facts established for disparity detection thresholds with narrowband stimuli:

- The disparity detection threshold is inversely proportional to the square root of contrast under a wide range of conditions (Legge and Gu, 1989). We shall refer to this relation as to *the square root law*.
- In dichoptic stimuli, disparity detection thresholds grow dramatically with the interocular difference in contrast (*the dichoptic contrast tuning function* for disparity detection; Halpern and Blake, 1988; Legge and Gu, 1989; Schor and Heckmann, 1989; see Fig. 6).

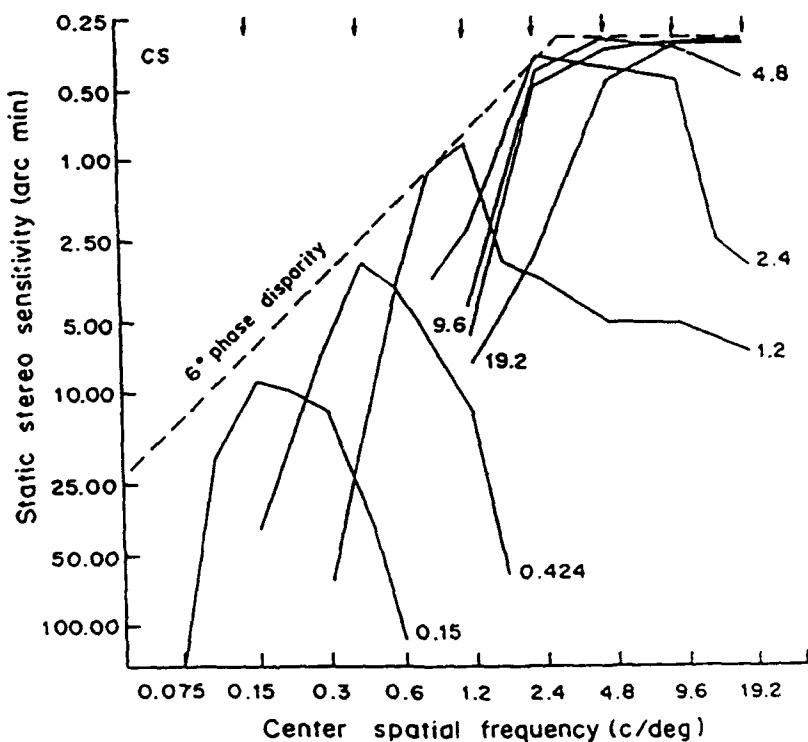


Fig. 7. Disparity detection thresholds (dashed line) and dichoptic tuning functions for stereopsis (solid curves), where dichoptic DoG profile stimuli have different center spatial frequencies. Invariant spatial frequency in spatial tuning estimates is shown by arrows (Schor, Wood and Ogawa, 1984, with permission).

c) Disparity sensitivity for narrowband stimuli increases with center spatial frequency below 2.5 cy/deg and is constant or declines again above 2.5 cy/deg Schor, Wood and Ogawa, 1984; Legge and Gu, 1989 (Fig. 7). There are also many other types of sensitivity that similarly change their behavior at the frequency of 2.5 cy/deg, for example, upper disparity limits and disparity thresholds for stereo motion and monocular motion (Schor and Wood, 1983; Schor, Wood and Ogawa, 1984; Westheimer, 1978; De Valois, 1982 (see Fig. 7.) It is important for our further analysis to note that *spatial tuning functions* obtained by measuring disparity detection threshold as a function of the ratio of peak frequencies of dichoptic stimuli also show, as demonstrated by Schor, Wood and Ogawa (1984), very different behavior above and below 2.5 cy/deg (see Fig. 7). If both parts of dichoptic stimulus were above 2.5 cy/deg, the thresholds were found to be virtually independent of the ratio between frequencies of the components. If, however, both parts were below 2.5 cy/deg, the thresholds grew rapidly with the center frequency difference.

d) The spatial sensitivity of stereo system was measured by Yang and Blake (1991) using monocular masking paradigm. Their data show masking sensitivity to peak at 5-8 cy/deg and to drop off rapidly at lower spatial frequencies. Any low-frequency stereo channels, if they exist, must have had negligible sensitivity to masks below 2 cy/deg.

Items c and d appear to be in contradiction, since the Schor, Wood and Ogawa results imply the existence of stereoscopic channels tuned as low as 0.15 cy/deg, whereas the Yang and Blake results restrict them to tunings above 2.5 cy/deg. We resolve this contradiction by developing a model in which there is no need to postulate any channels tuned to low frequencies; all low frequency effects can be explained by the effective contrast in the high-frequency channel most sensitive to the stimulus (i.e., the one with the lowest available center frequency).

The model is based on the following assumptions.

- i) All information for disparity processing is provided by narrowband channels.
- ii) All such channels are tuned to medium and high spatial frequencies with the lowest frequency channel peaking at about 2.5 cy/deg.
- iii) The channel tuned at 2.5 cy/deg has a symmetric balanced triphasic receptive field (such as a DoG spatial profile).
- iv) The precision of the disparity estimates is defined by the signals from the most sensitive channels at the corresponding points.

Consider a narrowband stimulus with an arbitrary center frequency below 2.5 cy/deg. The most sensitive channel for this stimulus will be the channel with the closest center frequency, which is the channel tuned to 2.5 cy/deg. Therefore, all narrowband stimuli at frequencies below 2.5 cy/deg will be processed solely by the channel tuned to 2.5 cy/deg.

As long as the operating channel and the stimuli are tuned to different frequencies, the output will constitute an effective contrast signal that depends on the discrepancy between the tuning frequencies of the channel and the stimulus\*. Based on the notion of a second-derivative operator we can derive a formula for effective contrast signal  $\chi$  for one-dimensional luminance profiles of the form

$$L(C, v, x) = L_0 \cdot (1 + C \cdot S(v, x)) \quad (4)$$

where  $S$  is a generic luminance profile of the family,  $C$  is the contrast,  $v$  is the spatial frequency and  $L_0$  is a background luminance. The second derivative of the profile can be expressed as:

$$L''(x) = L_0 \cdot C \cdot v^2 \cdot S''(v, x). \quad (5)$$

Without loss of generality, we may assume that the depth judgment is performed at the point  $x = 0$ . Then

\* The local effective contrast should not be taken as equivalent to perceived contrast. The perceived contrast is a result of comparison of luminances of possibly distant points. Such distant comparisons cannot be provided by a single filter: it requires long-range interactions as was shown by Land and McCann (1971).

Because the effective contrast ( $\chi$ ) is proportional to the second derivative (see above),

$$\chi \propto L'' \propto C \cdot v^2, \quad (7)$$

the effective contrast at 2.5 cy/deg channel is directly proportional to stimulus contrast but fits a quadratic law for spatial frequency.

The square root law for disparity detection thresholds  $\theta_d$  can be directly applied to the effective contrast signal. In its original formulation:

$$\theta_d \propto C^{-1/2}. \quad (8)$$

Now, when the spatial frequency is varied, it follows that

$$\theta_d \propto (\chi)^{-1/2} \propto C^{-1/2} \cdot v^{-1}. \quad (9)$$

Therefore, in the experiment where the contrast of the stimulus is kept constant, disparity detection thresholds must be reciprocal with spatial frequency. This is exactly the result in the study by Schor, Wood and Ogawa (1984). The slope of the disparity detection threshold for the DoG stimuli below 2.5 cy/deg was found to be equal to -1 (see Fig. 7) in accord with the present model.

Consider now the case where the left and right halves of a dichoptic stimulus excite the 2.5 cy/deg channel differently. The most direct way to introduce this difference is to present a low frequency dichoptic stimulus with different contrasts in the two half-images, which actually was the paradigm used

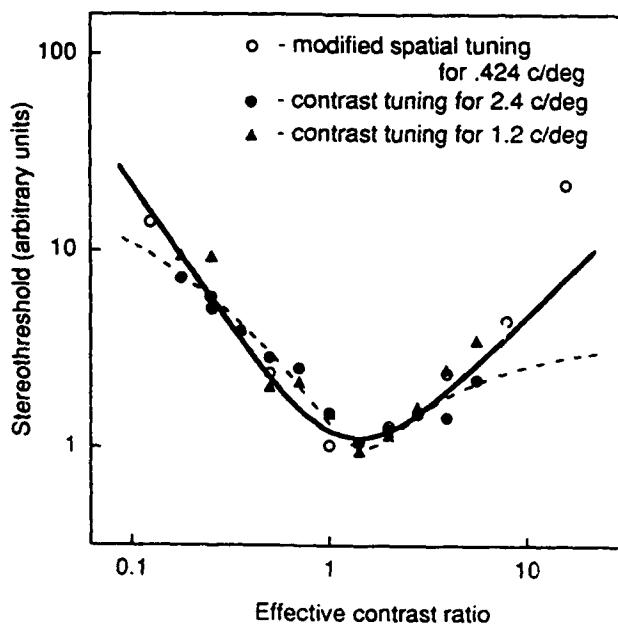


Fig. 8. Contrast tuning functions from Halpern and Blake (1988) (filled symbols) and a dichoptic spatial tuning function from Schor et al. (1984) (open symbols) where contrast and frequency ratios are converted into effective contrast ratios for channel tuned to 2.5 cy/deg. Upper curve is fit for 1.2 cy/deg data by our model (from Fig. 1); lower is best fit by Legge and Gu model. Vertical position is arbitrary.

by Halpern and Blake (1988). The measurements of the contrast tuning functions in that study for 1.2 and 2.4 cy/deg D10 narrowband stimuli should reveal, therefore, the contrast tuning function of the postulated 2.5 cy/deg channel, since our model predicts that all contrast tuning functions for stimuli below 2.5 cy/deg should have similar shape. The experimental data fit this prediction. The data for 4.8 cy/deg are different, however, because the stimulus is now falling in a range where more than one channel may be involved in the stereoscopic detection task..

A more ingenious way to stimulate the 2.5 cy/deg channel differently in the two eyes is to tune effective contrast by varying the frequency rather than the contrast of narrowband stimuli. This was the paradigm used by Schor, Wood and Ogawa (1984), where they presented a family of spatial tuning functions for stereopsis (Fig. 7). They measured disparity detection thresholds for dichoptic DoG stimuli of different widths. (When fused, DoGs of different widths produce a tilted surface, so the judgment of depth relative to fixation had to be performed at the central position of the fused image marked by fixation spot.)

According to assumption (iv), the spatial tuning functions and contrast tuning functions for stereopsis should be the same when spatial frequency and contrast variations both are expressed in terms of effective contrast. As long as effective contrast is affected by stimulus contrast linearly and by spatial frequency quadratically, the contrast tuning curves should be represented directly and the spatial tuning functions should be extended in horizontal direction by the factor of two (in log coordinates). The result of this transformation is presented in Fig. 8. Notice that all the corresponding curves have very similar shape, as is predicted by the theory.

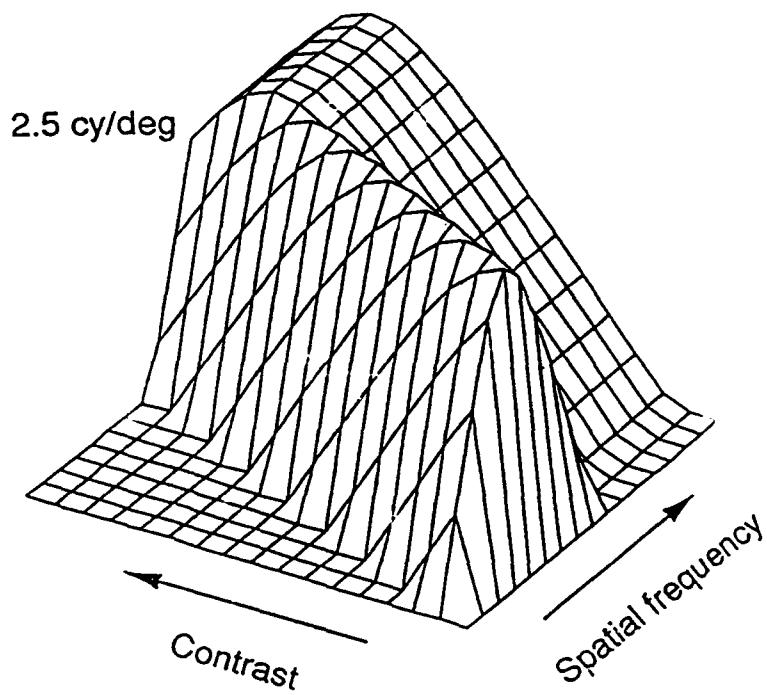


Fig. 9. Disparity detection threshold as a function of contrast and spatial frequency of one dichoptic stimulus while second is kept constant. Merge point of ridges depends on effective contrast of invariant stimulus. Model does not provide veridical prediction at high spatial frequencies: in fact, rear ridge should become flatter with spatial frequency increase.

### Control of Disparity Detection by Effective Contrast of the Dichoptic Stimuli

To summarize the analyses presented above we propose an explicit functional relation between disparity detection threshold behavior and stimuli parameters at low spatial frequencies. Replacing contrasts in the analytic description of contrast tuning by effective contrasts  $\chi_R$  and  $\chi_L$  for the right and left eyes we arrive at

$$\theta_d \propto \sqrt{\frac{(\chi_L)^p + k(\chi_R)^p}{(\chi_L)^{p+1}} + \frac{(\chi_R)^p + k(\chi_L)^p}{(\chi_R)^{p+1}}} \quad (10)$$

where  $p = 2$  and  $k = 0.8$ . This formula describes a behavior of disparity detection thresholds for the case where both dichoptic stimuli are tuned to spatial frequencies below 2.5 cy/deg. At present, we cannot expand this formula to high-frequency domain because we do not know the contrast interactions between dichoptic stimuli of different spatial frequencies. On qualitative level, assuming that binocular interaction does not depend on spatial frequency (while it depends) and that effective contrast for stimuli above 2.5 cy/deg does not depend on spatial frequency, our model predicts the stereo sensitivity surface presented in Fig. 9. There are two constraints on this surface:

- a) in the model we calculated the effective contrast by formula  $\chi = Cv^2 / 2.5^2$  which becomes slightly incorrect as the spatial frequency approaches to 2.5 cy/deg;
- b) the rear part of this surface corresponding to spatial frequencies above 2.5 cy/deg should flatten with increase of spatial frequency.

The main feature of this surface is that the ridge in low spatial frequency range has a slope of two from the top view. The cuts along the contrast axis are spatial tuning functions, the cuts along the spatial frequency axis which form the contrast tuning functions are two times wider.

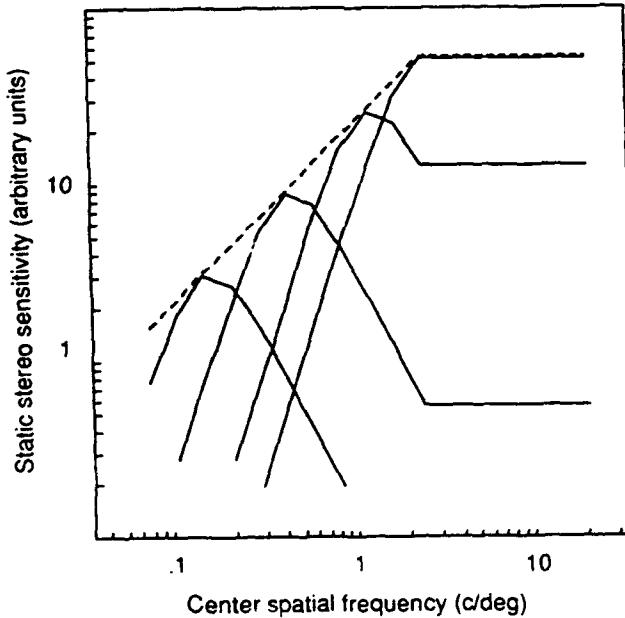


Fig. 10. Dichoptic spatial tuning functions for stereo predicted by the effective contrast model, contracted two times in horizontal direction to match contrast tuning function at 1.2 cy/deg fitted by our model for contrast tuning (upper curve from Fig. 1). Notice similarity with experimental data presented in Fig. 7.

An important consequence of our analysis is that the spatial tuning functions for low-frequency dichoptic stimuli must have the same shape, viz., the shape of the contrast tuning function of the 2.5 cy/deg channel contracted in log coordinates by a factor of two. Fig. 10 shows the contracted contrast tuning function, predicted by our model for contrast tuning, in the positions corresponding to the positions of spatial tuning functions in Fig. 3 (the relative vertical position of the curves along a constant phase disparity line is as predicted by the model). The similarity between the experimental data and the prediction of the model shown in Fig. 10 is impressive.

The analysis presented in this paper implies that stereo system does not use the signals from channels tuned below 2.5 cy/deg. Nevertheless, evidence for the existence of such channels for 2D spatial vision has been presented in several studies based on the masking paradigm (Stromeyer, Klein, Dawson and Spillmann, 1982; Wilson, MacFarlane and Phillips, 1983). We speculate that these low frequency channels, if they really exist, are not involved in stereoscopic processing. There are puzzling data for detection of trapezoidal stimuli presented by Campbell, Johnstone and Ross (1981) which call into question the involvement of low frequency bandpass channels in non-stereoscopic visual detection tasks.

Our conclusion that there are no channels with peak frequencies below 2.5 cy/deg involved in stereoscopic processing questions the principle underlying one of the most cited models of stereopsis (Marr and Poggio, 1979). This model holds that there is a range of stereomechanisms with different precision establishing correspondence between stereo images and that coarse low-frequency mechanisms provide initial data for finer high-frequency mechanisms. Our analysis suggests that this principle could apply only to the sub-range of relatively small disparities, in contradiction to the qualitative interpretation of the data of Schor et al. (1984) by the authors and by Tyler (1991) in terms of stereoscopic channels tuned to low spatial frequencies.

### High frequency stereoscopic channels

Having established that all the data below 2.5 cy/deg are compatible with the existence of a single second-derivative channel at that frequency, attention naturally turns to the question of the spatial structure of the stereoscopic channels above 2.5 cy/deg. For this question, the only relevant data are those reported by Yang and Blake (1991), who measured the masking effect on narrowband filtered stereograms of narrowband monocular noise, each with a variety of center frequencies of the noise bands. Monocular masking of narrowband stereoscopic depth targets produces masking sensitivity functions with very different tuning than for monocular detection. The peak masking frequencies were limited to the range of 3-8 c/deg, the peak frequency masker functions for higher test frequencies peaked at progressively lower mask frequencies. This unique behavior is a challenge to existing models of masking behavior for any perceptual domain.

Yang and Blake suggest that their masking sensitivity functions fall at two discrete frequencies and are compatible with a two-mechanism model of the spatial processing of disparity. However, they do not provide a full computational analysis to support this suggestion. Nevertheless, it is clear that the mechanisms of stereoscopic masking are very different from those of spatial contrast processing. We therefore developed a computational analysis of the stereoscopic masking data to determine how many mechanisms were required to provide a full account of the masking behavior, and to what extent they differed from spatial masking mechanisms.

### Model assumptions

Any computational analysis embodies a set of assumptions about the nature of the processing involved. For the present analysis, these assumptions are:

1. The spatial processing is implemented by a discrete set of mechanisms each selective to a different range of spatial frequencies. The spatial frequency tuning of each mechanism remains constant as test and mask stimuli are varied.
2. The channels are assumed to have the form of the power DoG described by equation 1, which has three free parameters for each channel: peak frequency, bandwidth (i.e., power) and sensitivity (assuming a fixed ratio between  $\sigma_1$  and  $\sigma_2$ ).

3. Sensitivity to the test stimulus is determined by probability summation between the separate mechanisms. In common with many other such models (Quick, 1974; Wilson et al., 1983; Lehky, 1985), the probability summation is modeled as a non-Euclidean vector summation process.
4. The masking process is assumed to be characterized by a multiplicative masking amplitude function that is invariant with spatial frequency of test or mask (eq. 2). Each mechanism is assumed to become desensitized by the mask in proportion to its tuning sensitivity at the mask frequency scaled in strength by the masking amplitude function.
5. The masking amplitude function was assumed to be identical for all channels. This assumption was made in order to limit the number of free parameters for this initial evaluation of the channel structure of stereopsis. The assumption is known to be violated for spatial adaptation (Blakemore and Campbell, 1970) and masking (Wilson, McFarlane and Phillips, 1983), but may provide a useful first approximation for a more limited system such as stereopsis.

### Model implementation

The model was implemented computationally by a set of procedures to allow practical realization of the basic concepts. The basic stereoscopic model was of multiple independent spatial channel with probability summation, as described by eqs. 1-3. Nevertheless, the implementation differed substantially from the typical approach of Wilson, McFarlane and Phillips (1983), for example. Rather than maintaining a constant mask strength, Yang and Blake (1991) employed the more powerful procedure of presenting a fixed test stimulus and varying the mask strength to threshold visibility for the test stimulus. This procedure, which is akin to that of Stiles (1939), has the property of revealing segments of the underlying channels uncontaminated their neighbors.

To derive the expression for the threshold masking function for a system of channels of sensitivity  $S_i$  at test and mask frequencies  $f_t$  and  $f_m$ , we assume that the internal response at threshold  $\Delta R$  is given by the product of the sensitivity to the test frequency and the masking effect  $M_i$  at the masking frequency for each channel  $i$

$$\Delta R_i = a_t S_i(f_t) \cdot M_i(a_m S_i(f_m)) \quad (11)$$

where  $a_t$  and  $a_m$  are scaling constants. The Yang and Blake paradigm is such that  $a_m$  is the dependent variable, which may be obtained through the inverse of the masking amplitude function  $M^{-1}$ , if it is assumed to be equal for all channels, when the threshold masking behavior for each channel acting alone would be

$$A_{m,i} = \frac{M^{-1}\left(\frac{\Delta R}{a_t S_i(f_t)}\right)}{S_i(f_t)} \quad (12)$$

If we assume that the internal response is masked proportionately to some power  $q$  of the mask amplitude

$$R \propto a^{-q} \quad (13)$$

then

$$M^{-1} \propto R_i^{-\frac{1}{q}} \quad (14)$$

giving rise to the masking equation used for the analysis of the threshold that would be predicted for each channel alone

$$A_{m,i} \propto \frac{a_i S_i(f_i)^{-\frac{1}{q}}}{S_i(f_m)} \quad (15)$$

which are then combined to give the overall threshold for the system by a winner-take-all rule

$$A_m = \max_i(A_{m,i}) \quad (16)$$

The consequence of this analysis is that every masking sensitivity function consists of the intersection of the *least* sensitive segments of any channel at each mask position (because it is these that are the least masked, and therefore the most sensitive under the masking conditions).

The model was computed in the Matlab matrix algebra language on a Macintosh Quadra. It was fitted to the data through back-propagation of the mean square error by means of a steepest descent algorithm in a  $3K + 3$  parameter space, where  $K$  was the number of channels. At each iteration, the algorithm calculated the parameter-normalized unit vector to determine the optimum weighting for error minimization in the local region of the parameter space. A large step was then made in the direction of the unit vector, decreasing progressively by steps of 0.5 until the minimum error was achieved.

The Yang and Blake masking data at 8 test and 8 masking spatial frequencies formed the measured masking surface (the matrix was incomplete where sensitivities were below measurable values, resulting in ~50 free parameters in the data). The models fitting these data varied from  $N = 2$  mechanisms (9 parameters) to  $N = 8$  mechanisms (27 parameters). For the present analysis, the analysis was limited to the data for the finest disparity condition, 1.3' disparity for the central square relative to its surround. The reason for this is that at the larger disparities (8' and 15') the monocular noise in the highest center

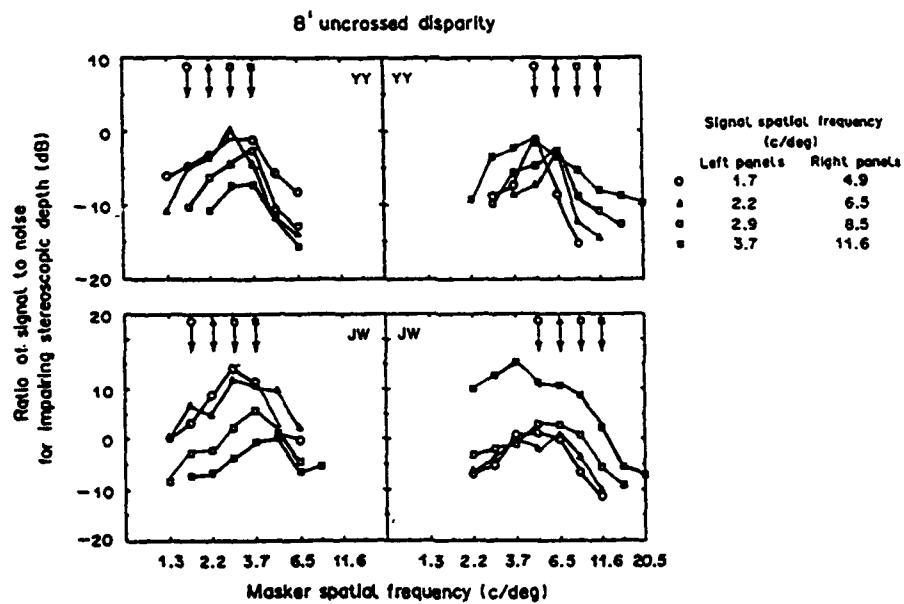


Fig. 11. Effect of monocular masking by 2D narrowband noise on detectability of narrowband stereograms of 1.3' disparity for two observers (Yang and Blake', 1991).

frequency conditions might have interacted with the grain in the disparate regions to produce spurious signals at disparities lower than the computed disparity of the center square, giving a spurious indication of depth discrimination under conditions where the signal might otherwise have been masked (if the noise were binocular, for example). Thus, the larger disparity conditions require a more sophisticated analysis than has been developed so far.

The masking data for 1.3' disparity are shown in Fig. 11A&B for Yang and Blake's two observers. The masking sensitivity functions vary dramatically in strength over the range of peak test frequencies, showing the curious behavior at high frequencies for the channel bandwidths to become broader at the highest frequencies. This is the masking data base from which we will evaluate the channel structure of fine stereopsis.

The optimum model fits for two mechanisms are shown in Fig. 12, where panels A&B show the unmasked sensitivities of the two mechanisms and C&D the fits to the masking data. The fits are accurate to a residual variation of 0.16 log unit for observer YY and 0.35 log unit for observer JW, thus accounting for a high proportion of the 2 log unit range of variation present in the data. (Remarkably, there was no significant tendency for the error to decrease as the number of channels in the model was increased from 2 to 8. The errors appeared to be randomly distributed about a fixed value over this range.) In some respects, the fits are reminiscent of the two-channel model proposed by Yang and Blake, with one channel peaking at about 2.5 cy/deg and the other at about 7 cy/deg. Note, however, the substantial variation in estimated width of the lower channel required by the data that it dominates at low spatial frequencies, the masking sensitivity functions being narrow for YY and broad for JW.

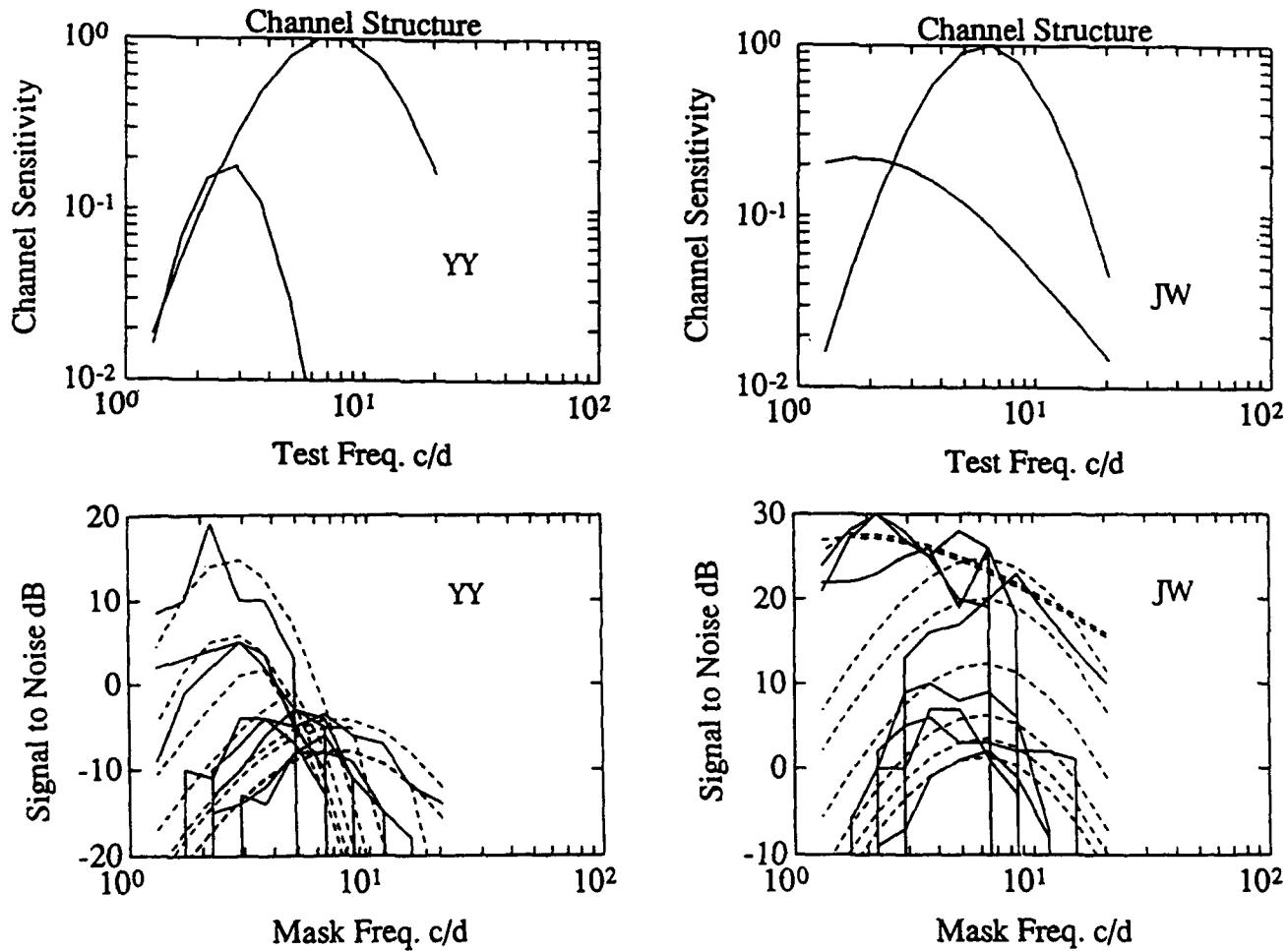


Fig. 12. Optimum model fits to the data of Fig. 11.

Nevertheless, the two-channel model is incomplete in two respects. First, the predicted peak positions of its masking sensitivity functions fail to capture the variations in masking behavior at high frequencies present in the data (Fig. 12 A&B), since the property of dominance by the least sensitive channel implies that the predicted masking amplitude functions are all of identical form at high frequencies. We anticipate that this mismatch will be improved by the use of different masking amplitude functions for each channel, since the channel shape is determined as much by the variation across test frequencies as by the shapes of the masking sensitivity functions.

Second, the proposal by Yang and Blake that the peak sensitivities of the masking sensitivity functions represent the unmasked sensitivity of the underlying channels is incorrect. Fig. 13 shows that the peak masking sensitivity of the model provides a good representation of the data, but that the unmasked sensitivity of the channels that generated these data is quite different. This result underlines the implication of the present analysis that only way to determine any of the features of the mechanisms underlying masking behavior of the type described by Yang and Blake is to perform a computational optimization.

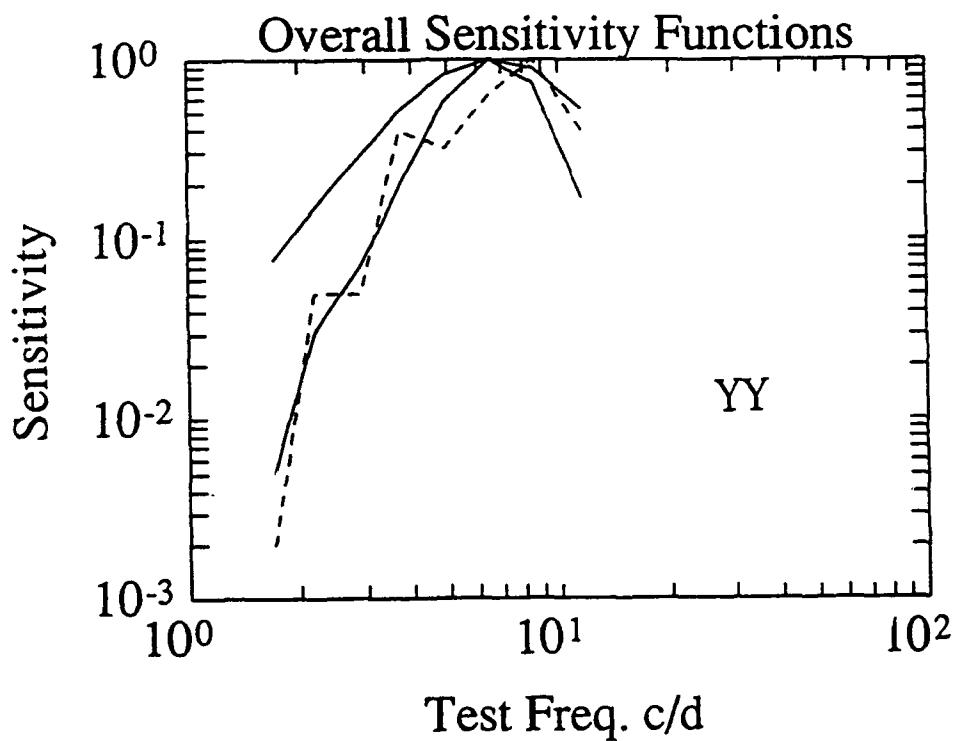


Fig. 13. Peak masking sensitivity as a function of test frequency from the data of Fig. 11A and as predicted from the model fits (lower points and curve). Note that these functions deviate markedly from the overall unmasked sensitivity function of the model (upper curve).

## REFERENCES

Anderson S.J. and Burr D.C. (1985) Spatial and temporal selectivity of the human motion detection system. *Vision Research* **25**, 1147-1154.

Anderson S.J. and Burr D.C. (1989) Receptive field properties of human motion detector units inferred from spatial frequency masking. *Vision Research* **29**, 1343-1358.

Beverley K.I. and Regan D. (1973) Evidence for the existence of neural mechanisms selectively sensitive to the direction of movement in space. *Journal of Physiology* **235**, 17-29.

Blakemore C. and Campbell F.W. (1969) On the existence of neurons in the visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology* **203**, 237-260.

Cameron, E.L., Baker C.L. & Boulton J.C. (1992) Spatial frequency selective mechanisms underlying the motion aftereffect. *Vision Research* **32**, 561-568.

Campbell, F. W., Johnstone, J. R. and Ross, J. (1981) An explanation for the visibility of low frequency gratings. *Vision Research* **21**, 723-730.

De Valois, R. (1982) Early visual processing: feature detection or spatial filtering? in (Albrecht, D., Ed.), *Recognition of Pattern and Form*, pp. 152-172. Springer, Berlin.

Georgesom M.A. and Harris M.G. (1984) Spatial selectivity of contrast adaptation: Models and data. *Vision Research* **24**, 729-742.

Halpern, D. L. and Blake, R. (1988) How contrast affects stereoaucuity. *Perception* **17**, 483-495.

Hess R.F. and Snowden R.J. (1992) Temporal properties of human visual filters: Number, shapes and spatial covariation. *Vision Research* **32**, 47-59.

Legge G. E. (1981) A power law for contrast discrimination. *Vision Research* **21**, 457-469.

Legge, G. E. and Gu, Y. (1989) Stereopsis and contrast. *Vision Research* **29**, 989-1004.

Lehky S. (1985) Temporal properties of visual channels measured by masking. *Journal of the Optical Society of America A2*, 1260-1272.

Quick R.F. (1974) A vector magnitude model of contrast detection. *Kybernetik* **16**, 65-67.

Schor, C. M. and Heckmann, T. (1989) Interocular differences in contrast and spatial frequency: effects on stereopsis and fusion. *Vision Research* **29**, 837-847.

Schor, C. M., Wood, I. C. (1983) Disparity range for local stereopsis as a function of luminance spatial frequency. *Vision Research* **23**, 1649-1654.

Schor, C. M., Wood, I. C. and Ogawa, J. (1984) Spatial tuning of static and dynamic local stereopsis. *Vision Research* **24**, 573-578.

Stiles W.S. (1939) The directional sensitivity of the retina and the spectral sensitivities of the rods and cones. *Proceedings of the Royal Society, London B* **127**, 64-105.

Stromeyer C.F. and Klein S.A. (1974) Spatial frequency channels in human vision as asymmetric edge mechanisms. *Vision Research* **14**, 1409-1420.

Stromeyer, C. F. III, Klein, S., Dawson, B. M. and Spillmann, L. (1982) Low spatial-frequency channels in human visual: adaptation and masking. *Vision Research* **22**, 225-233.

Westheimer, G. (1978) Spatial phase sensitivity for sinusoidal grating targets. *Vision Research* **18**, 1073-1074.

Wilson H.R. and Bergen J. (1979) A four-mechanism model for threshold spatial vision. *Vision Research* **19**, 19-32.

Wilson, H. R., MacFarlane, D. K. and Phillips, G. C. (1983) Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research* **23**, 873-882.

Yang, Y. and Blake, R. (1991). Spatial frequency tuning of human stereopsis. *Vision Research* **31**, 1177-1189.

Supported by NIH grant EY 7890.

Reconstruction of 3D structure  
from two central projections by a fully linear algorithm

Maxim L. Kontsevich<sup>1</sup> and Leonid L. Kontsevich<sup>2</sup>

<sup>1</sup> Max-Plank Institut fuer Mathematik, Bonn

<sup>2</sup> Smith-Kettlewell Eye Research Institute, San Francisco

ABSTRACT

A completely linear method for reconstruction of 3D structure from two central projections is proposed. We show that this method for combining such parameters as speed, robustness and generality performs better than all other algorithms.

1. INTRODUCTION

In his classical book on motion, Ullman<sup>1</sup> showed that, for central projections, five point correspondences over two views are sufficient to determine the structure of the point-rigid configuration. The first algorithmic solutions proposed by Ullman<sup>1</sup> and by Roach and Aggarwal<sup>2</sup> were computationally ineffective mainly because they were nonlinear. A real breakthrough was made by Longuet-Higgins<sup>3</sup>, who proposed an algorithm that provided a linear solution for non-linear combination of parameters. His approach was further developed by Tsai and Huang<sup>4</sup>, Faugeras et al.<sup>5</sup>, and Weng et al.<sup>6</sup>. However, the algorithms of this class are particularly sensitive to input errors because the linear constraint they utilize is not the same as the rigidity constraint.

The tolerance to input errors was dramatically improved by Heeger and Jepson<sup>7</sup> for the case of instantaneous-time formulation. However, in this formulation a vector field of velocities is available instead of two arbitrary views that are required in standard and more general discrete-time formulation. The method initially proposed by Heeger and Jepson is imperfect from the computational point of view: it uses large matrices at all stages of processing and has a time-consuming minimization of residual function for reconstruction of translation direction (this stage was slightly improved recently<sup>8</sup>).

In this paper we present a new method which is completely linear, robust and more computationally effective than others. In contrast with the Jepson-Heeger method, our method is designed for discrete-time formulation. This method was initially published in Kontsevich et al.<sup>9</sup> where we proposed a general approach to 3D structure reconstruction for the cases of orthographic, weak perspective, and central projections. Because this paper is not available in English and because L.Kontsevich<sup>10</sup> later designed a more advanced algorithm for the weak perspective case, we reproduce here a part of the above-mentioned paper related to the central projection case and present a comparison of the performance between the Jepson-Heeger and our methods.

## 2. THE PROBLEM

Let us pose a general problem of depth reconstruction from several projections. Assume that in three-dimensional Euclidean space  $S$  the finite set of points  $A = \{a_i\}$  is selected (let  $n$  is the number of points). This set we call *object*. Assume that on  $S$   $m$  projection operators  $p_j$  on two-dimensional Euclidean spaces  $F_j$  are designed. In each  $F_j$  for point  $a_i \in A$  corresponds point  $a_{ji} = p_j(a_i)$ .

Assume that neither  $A \subset S$  nor  $p_j$  are known. We suppose that the correspondence between points  $a_{ji}$  on projections is established (i.e., it is known which points are projections of the same point of the object). The class of possible mappings  $p_j$  is also supposed to be known. The problem is to reconstruct  $A$ .

We restrict our task here to the case of two central projections with known parameters of the optic system (focus distance, to be precise).

## 3. THE SOLUTION

### 3.1. Object reconstruction up to projective transformation

Let  $V$  be an affine space. Denote by  $\tilde{V}$  a corresponding vector space with  $\dim \tilde{V} = \dim V + 1$  where  $V$  is identified with a hyperplane that does not pass through zero. Projective space  $P(\tilde{V})$  consists of points of  $V$  and infinitely distant points. A mapping of central projection  $p_j: S \rightarrow F_j$  can be raised up to multiplier to some linear mapping

$\tilde{p}_j: \tilde{S} \rightarrow \tilde{F}_j$ . Notice that to points  $a_i$  in space  $S$  and  $a_{ji}$  in  $F_j$ , will correspond one-dimensional spaces  $\tilde{a}_i$  and  $\tilde{a}_{ji}$  in  $\tilde{S}$  and  $\tilde{F}_j$ , respectively. Then,  $\tilde{p}_j: \tilde{a}_i \rightarrow \tilde{a}_{ji}$ .

Consider a mapping  $\tilde{p} = \tilde{p}_1 \oplus \tilde{p}_2: \tilde{S} \rightarrow \tilde{F}_1 \oplus \tilde{F}_2$ . In the case where the centers of projections do not coincide  $\text{Ker}(\tilde{p}) = \text{Ker}(\tilde{p}_1) \cap \text{Ker}(\tilde{p}_2) = \tilde{f}_1 \cap \tilde{f}_2 = 0$  (here  $\tilde{f}_j$  is the center of  $j$ -th projection). Therefore,  $\tilde{S}$  can be identified as linear space with  $\tilde{p}(\tilde{S})$ . So, the reconstruction of projective structure of the object can be reduced to finding 4-dimensional subspace  $\tilde{S}$  in 6-dimensional space  $\tilde{F}_1 \oplus \tilde{F}_2$ , for which  $\dim(\tilde{S} \cap (\tilde{a}_{1i} \oplus \tilde{a}_{2i})) \geq 1$  for any  $i$ . This condition is equivalent to the existence of the point  $a_i$  projecting in  $a_{1i}$  and  $a_{2i}$ . We shall show below that the mentioned constraints on  $\tilde{S}$  in the case of sufficient number of points in  $A$  define a one-parameter family of possible solutions equivalent from the point of view of projective structure. In concrete calculations it is sufficient to get only one of them.

As long as the codimension of  $\tilde{S}$  is equal to two, it is reciprocal to decomposable 2-form  $\omega \in \wedge^2((\tilde{F}_1 \oplus \tilde{F}_2)^*)$  defined up to scalar multiplier. Let us denote by  $T$  a component of  $\omega$  in the term  $\tilde{F}_1^* \otimes \tilde{F}_2^*$  from the decomposition in the direct sum  $\wedge^2((\tilde{F}_1 \oplus \tilde{F}_2)^*) = \wedge^2(\tilde{F}_1^*) \oplus (\tilde{F}_1^* \otimes \tilde{F}_2^*) \oplus \wedge^2(\tilde{F}_2^*)$ . It can be shown that tensor  $T$  defines the space  $\tilde{S}$  up to the action of independent dilatations in  $\tilde{F}_1$  and  $\tilde{F}_2$ . The constraints on  $\tilde{S}$  can be rewritten in the following form:

$$\langle T, \tilde{a}_{1i} \wedge \tilde{a}_{2i} \rangle = \langle \omega, \tilde{a}_{1i} \wedge \tilde{a}_{2i} \rangle = 0.$$

Let us give the constraints on  $T$  in explicit form. The subspace is defined by two linear equations:

$$(\alpha_1^T, \alpha_2^T)(v_1, v_2)^T = 0; \\ (\beta_1^T, \beta_2^T)(v_1, v_2)^T = 0.$$

Here  $v_j \in \tilde{F}_j$ ;  $\alpha_j^T, \beta_j^T \in \tilde{F}_j^*$ . In matrix form  $T = \beta_2 \alpha_1^T - \alpha_2 \beta_1^T$ . Constraints on the subspace  $\tilde{S}$  mean that

$$\det \begin{pmatrix} \alpha_1^T \tilde{a}_{1i} & \alpha_2^T \tilde{a}_{2i} \\ \beta_1^T \tilde{a}_{1i} & \beta_2^T \tilde{a}_{2i} \end{pmatrix} = 0 = (\beta_2^T \tilde{a}_{2i})(\alpha_1^T \tilde{a}_{1i}) - (\alpha_2^T \tilde{a}_{2i})(\beta_1^T \tilde{a}_{1i}) = \\ tr((\tilde{a}_{2i}^T \beta_2 \alpha_1^T \tilde{a}_{1i}) - (\tilde{a}_{2i}^T \alpha_2 \beta_1^T \tilde{a}_{1i})) = tr((\beta_2 \alpha_1^T - \alpha_2 \beta_1^T)(\tilde{a}_{1i} \tilde{a}_{2i}^T)) = tr(T(\tilde{a}_{1i} \tilde{a}_{2i}^T)).$$

Thus, we obtained linear equations defining tensor  $T$ . As long as  $\text{rank}(T)=2$ ,  $T$  can be represented as a difference of two matrices of rank 1:  $T = T_1 - T_2$ ,  $\text{rank}(T_1) = \text{rank}(T_2) = 1$ . Decomposing  $T_1$  and  $T_2$  into the product of one column and one row vectors we obtain the coefficients of equations defining one of the possible subspaces  $\tilde{S}$  (it was mentioned above that the solution of this task is not unique).

### 3.2. Reconstruction of the Euclidean structure of the object

Let us introduce non-degenerated scalar product on  $\tilde{F}_j$  in the following manner. As long as  $F_j$  is defined by an optic system with known focus distance,  $\tilde{F}_j$  can be canonically identified with three-dimensional physical space where the focus distance can be naturally considered as a distance unit.

Let us show that Euclidean structure (up to scalar multiplier) in  $S$  defines a quadratic form  $Q^*$  with signature (+ + + 0) in  $\tilde{S}^*$  (up to multiplier) and vice versa. First of all,  $\tilde{S}$  contains a linear subspace  $L$  which is parallel to  $S$  and, therefore, possesses by non-degenerated quadratic form. Then, the scalar product is defined in  $L^* = \tilde{S}^* / \text{Ann}(L)$  and, therefore, in  $\tilde{S}^*$ . In backward direction, the quadratic form  $Q^*$  of the noted signature possesses a one-dimensional kernel whose annulator is a hyperplane  $L$  in  $\tilde{S}$ . The metrics of  $L$  can be translated on any hyperplane parallel to  $L$  and, therefore, up to multiplier on the part  $P(\tilde{S}) \setminus P(L)$  of projective space. Then  $\tilde{p}_j^* : \tilde{F}_j^* \rightarrow \tilde{S}^*$  are isometric embeddings up to multiplier.

The space  $M = \tilde{p}_1^*(\tilde{F}_1^*) \cap \tilde{p}_2^*(\tilde{F}_2^*)$  has dimension two, and two quadratic forms arising from Euclidean structures in  $\tilde{F}_1^*$  and  $\tilde{F}_2^*$  must be proportional. Let us multiply one of the quadratic forms in  $\tilde{F}_j^*$  by an appropriate coefficient such that those constraints in  $M$  would be the same. Let us select orthonormal basis  $\{e_1^*, e_2^*\}$  in  $M$ . Denote by  $e_3^*$  and  $e_4^*$  vectors from  $\tilde{p}_1^*(\tilde{F}_1^*)$  and  $\tilde{p}_2^*(\tilde{F}_2^*)$  respectively such that  $\{e_1^*, e_2^*, e_3^*\}$  and  $\{e_1^*, e_2^*, e_4^*\}$  are

orthonormal bases in the corresponding subspaces. The Gram matrix of the quadratic form  $Q^*$  in the basis  $\{e_1^*, e_2^*, e_3^*, e_4^*\}$  is

$$Q^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \lambda \\ 0 & 0 & \lambda & 1 \end{pmatrix}$$

where  $\lambda$  is unknown value. From the constraint  $\det(Q^*)=0$  we immediately obtain  $\lambda=\pm 1$ . Selecting one of these values of  $\lambda$  we find two solutions.

Let us associate the point from  $p(S)$  with coordinates  $(x_1, x_2, x_3, x_4)$  in the basis  $e_k$  with the point in  $R^3$  with coordinates  $x_1 / (x_3 - \lambda x_4), x_2 / (x_3 - \lambda x_4), x_3 / (x_3 - \lambda x_4)$ . The constructed mapping of  $p(S)$  in  $R^3$  is isometric.

#### 4. THE ALGORITHM

1. Let us introduce the orthogonal system on the projection  $F_j$ . Assume that the origin will be the point on the projection which is the closest most close to the center of the projection. Let the unit of length in the coordinate system linked with  $j$ -th projection be equal to the focus distance. Denote the coordinates of the points  $a_i$  in the constructed coordinate system by  $x_{ij}$  and  $y_{ij}$ .

2. Matrix  $T$  can be found from the system of linear equations:

$$\operatorname{tr} \left( T \begin{pmatrix} x_{1i}x_{2i} & x_{1i}y_{2i} & x_{1i} \\ y_{1i}x_{2i} & y_{1i}y_{2i} & y_{1i} \\ x_{2i} & y_{2i} & 1 \end{pmatrix} \right) = 0.$$

This is a homogeneous system of linear equations with 9 unknowns. Non-zero solution can be found up to multiplier from 8 points.

This overdefined system of homogeneous linear equations can be solved by the standard methods of linear algebra.

This pseudoinversion algorithm allows to reconstruct  $T$  with precision  $\propto n^{-1/2}$  ( $n$  is a number of points of object) while the variance of errors of point coordinates are invariant. Matrix  $T$  is defined by mutual arrangement of projection planes and centers, and vice versa, it defines it up to scaling.  $T$  can be reconstructed with significantly higher precision than the shape of the object.

If the calculated matrix  $T$  has close to zero determinant, the calculations can be continued. Otherwise the projections are incompatible.

3. Let us arbitrarily decompose  $T$  in the difference of rank-1 matrices  $T_1$  and  $T_2$ . For example, for this purpose the standard procedure of singular decomposition can be exploited. The rank-1 matrices  $T_1$  and  $T_2$  should be arbitrarily represented as a products of column and raw:

$$T = T_1 - T_2 = \beta_2 \alpha_1^T - \alpha_2 \beta_1^T.$$

Thus, we obtain the columns of coefficients  $\alpha_1$ ,  $\alpha_2$ ,  $\beta_1$  and  $\beta_2$ . As long the singular decomposition is used,  $\alpha_1^T \beta_1 = \alpha_2^T \beta_2 = 0$  and  $\alpha_1^T \alpha_1 = \beta_1^T \beta_1 = 1$ . The condition for consistency of quadratic forms means that  $\alpha_2^T \alpha_2 = \beta_2^T \beta_2 = c^2$ . This condition must be checked: if it fails, the task does not have a solution.

4. The basis in the space  $\tilde{S}^*$  which can be considered as a factor space

$$R^6 / \left( * \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + * \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \right)$$

is the following:

$$e_1^* = \begin{pmatrix} \alpha_1 \\ 0 \end{pmatrix}, e_2^* = \begin{pmatrix} \beta_1 \\ 0 \end{pmatrix}, e_3^* = \begin{pmatrix} \alpha_1 \times \beta_1 \\ 0 \end{pmatrix}, e_4^* = \begin{pmatrix} 0 \\ \alpha_2 \times \beta_2 \end{pmatrix} / c$$

5. Now one can construct a representative  $v_i \in \tilde{S} \subset R^6$  for each point  $a_i$ :  $v_i = c_1(x_{1i}, y_{1i}, 1, 0, 0, 0)^T + c_2(0, 0, 0, x_{2i}, y_{2i}, 1)^T$ . The coefficients  $c_1$  and  $c_2$  must be chosen in such a way that  $(\alpha_1^T, \alpha_2^T) \cdot v_i = 0$ . For example,

$$c_1 = \alpha_2^T(x_{2i}, y_{2i}, 1)^T, c_2 = -\alpha_1^T(x_{1i}, y_{1i}, 1)^T$$

6. Now for each point  $a_i$  its four coordinates  $x_k(a_i) = e_k^* v_i$  can be calculated. Then, two Euclidean models can be calculated (for  $\lambda = 1$  and for  $\lambda = -1$ ) by the presented earlier formula

$$x_1 / (x_3 - \lambda x_4), x_2 / (x_3 - \lambda x_4), x_3 / (x_3 - \lambda x_4).$$

One of these models corresponds to similar orientation of projection centers (both between object and projection plane or both behind the plane) and another corresponds to opposite orientations. Under natural conditions the latter model is meaningless. To recognize which model is correct the signs of coefficients  $c_1$  and  $c_2$  have to be overviewed. If for all points these coefficients have opposite signs,  $\lambda = -1$  should be used. If the signs for each point are the same,  $\lambda = 1$ . The situation where the coefficient for some points have the same signs and for some different is impossible: it means that the projection plane intersects the object.

### 5. COMPARISON WITH THE JEPSON-HEEGER METHOD

To compare our method with that of Heeger and Jepson<sup>7</sup> we implemented their algorithm. They did not describe in detail how they integrated the solutions from the regions into which the visual field was subdivided; for this reason we did not use this subdivision. This modification led to increased precision of structure reconstruction but reduced speed.

The input objects were computer-generated. The object consisted of 20 points randomly chosen within the cube with the side equal to 100 units and with the center at the origin. The projection center had coordinates (0, 0, 150) with projection plane parallel to the plane XY and unity focus distance.

In computer experiments, the object performed the designed motion and the projections for analysis were obtained before the motion and after its completion. The rotational component of the motion was always around the  $Y$  axis. After exact projections were obtained, the independent noise to the point positions on both projections was added. The noise had the same variance along both dimensions in the plane and was evenly distributed. The numbers presented on the graphs represent the width of the noise range in one dimension.

To estimate the performance of the algorithms we used the following procedure. As long as 3D structure reconstruction from central projections can be done up to scaling the length ratios for all corresponding edges should be the same. We considered a chain of edges that passed once through each point of the object by connecting the points in the order of their generation by the computer. For each segment from this chain in the original and reconstructed objects the lengths were calculated. Then we scaled the reconstructed object to minimize the median of the absolute value of the difference between original and reconstructed lengths of the edges from the chain. This median was used as an estimate of the depth reconstruction. This procedure was performed 21 times for each set of parameters and the resulting estimate was a median of 21 estimates. The use median instead of least squares was because of instability of depth estimates at high levels of noise.

In Fig. 1a the error as a function of noise magnitude is presented for the translation vector (1, 2, 3), which is small relative to the object size. The data show that our method provides precise 3D reconstruction for no noise case while the Jepson-Heeger method produces some error due to imperfection of instantaneous approximation. At larger noise levels Jepson-Heeger method provides slightly better estimates than ours.

In Fig. 1b we demonstrate that, while the precision of the Jepson-Heeger method is good for small movements, the range over which it works stably is actually narrow, and in the majority of the cases our method works better. In this computer experiment the rotation angle was varied while the noise amplitude and the translational component were constants equal to 0.001 and (1, 2, 3), correspondingly. The Jepson-Heeger method fails at a rotation angle as small as 4 degrees, while our method dramatically exceeds its performance for larger rotation angles.

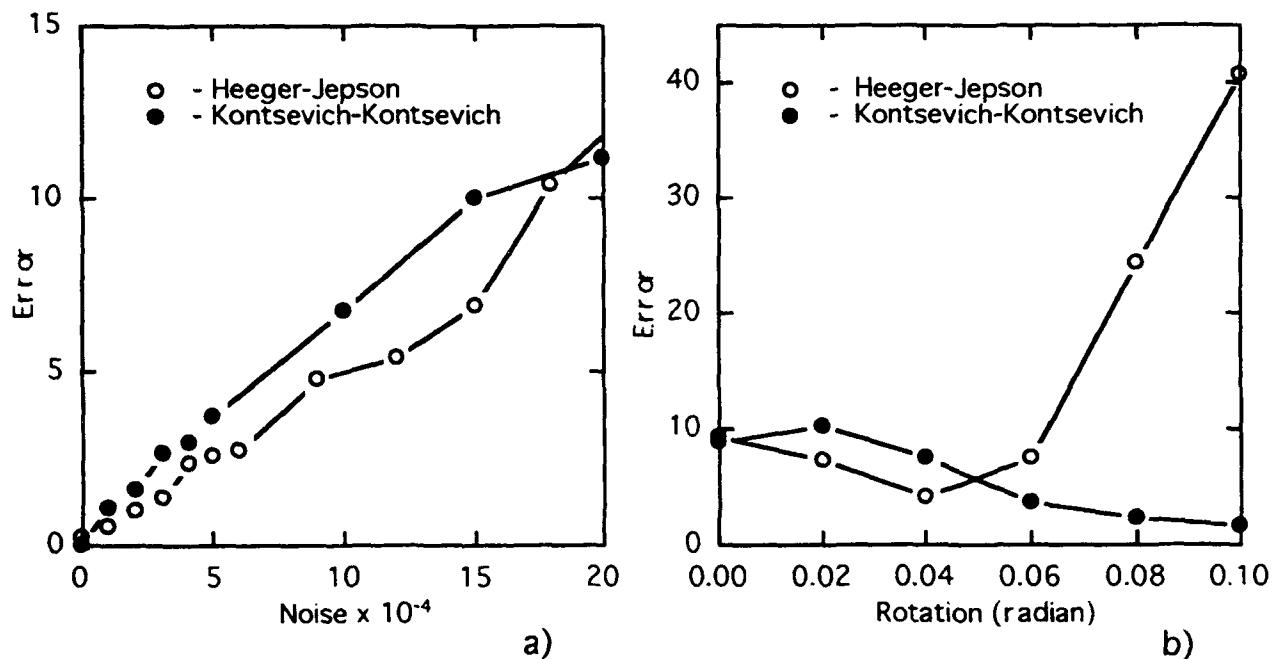


Fig. 1. Comparison of the Jepson-Heeger and Kontsevich-Kontsevich methods. a) precision of the algorithms at different noise levels; b) precision of the algorithms for different rotation angles. Performance across the noise is similar with both algorithms. For rotations, however, performance of the Kontsevich-Kontsevich algorithm improves with rotation angle, while the Jepson-Heeger method fails at the angle of 4 degrees.

We do not provide a detailed comparison of the speed for Jepson-Heeger and our methods because performance for either method strongly depends on the specifics of implementation. However, compared to our probably-imperfect implementation of the Jepson-Heeger method, the speed of our method was at least two orders higher. And, theoretically, this difference must increase with the number of points. Reason for this dramatic difference is that, while the Jepson-Heeger method works with matrices whose size is proportional to the number of points and employs a computationally expansive search for residual function for reconstruction of translation direction, our method uses fast standard procedures from linear algebra for matrices of small dimension.

## 6. CONCLUSIONS

We propose for reconstruction of 3D structure from two central projections a new linear method that is better than all other methods except the method proposed by Heeger and Jepson. The latter method performs slightly better than our performance for small

movements, but it is not applicable for large movements, for which the precision of our method increases. Our method is faster than the Heeger and Jepson method, and the memory allocated for parameters is independent of the number of points. These features make this method the first computationally effective robust method for 3D structure reconstruction from central projections in discrete-time formulation.

#### 7. ACKNOWLEDGMENT

This study was partly supported NIH Grant EY 7890 for L. Kontsevich.

#### 8. REFERENCES

1. S. Ullman. The Interpretation of Visual Motion. (MIT Press, Cambridge, Mass., 1979).
2. J. W. Roach and J. K. Aggarwal, "Determining the movement of objects from sequence of images." IEEE Trans. Patt. Anal. Mach. Intell. PAMI-2, 554-562 (1980).
3. H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections." Nature 293, 133-135 (1981).
4. R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces." IEEE Trans. Patt. Anal. Mach. Intell. PAMI-6, 13-27 (1984).
5. O. D. Faugeras, F. Lustman and G. Toscani, "Motion and structure from motion from point and line matches." Proc. 1st Intern. Conf. Comput. Vision, London, June, 25-34 (1987).
6. J. Weng, T. S. Huang and N. Ahuja, "Motion and structure from two perspective views: Algorithms, error analysis, and error estimation." IEEE Trans. Patt. Anal. Mach. Intell. PAMI-11, 451-476 (1989).
7. D. J. Heeger and A. D. Jepson, "Subspace methods for recovering rigid motion I: algorithm and implementation." Intern. J. Comp. Vision 7, 95-117 (1992).
8. A. D. Jepson and D. J. Heeger, "Linear subspace methods for recovering translational direction." To appear in Spatial Vision in Humans and Robots, Harris and Jenkin eds., Cambridge University Press, 1992.
9. L. L. Kontsevich, M. L. Kontsevich, and A. H. Shen, "Two algorithms of shape recovery." Autometria No 5, 72-77 (1987). [in Russian]
10. L. L. Kontsevich, "The inference of three-dimensional structure from weak perspective projections based on pairwise comparisons of images." Proc. SPIE 1669, 51-60 (1992).

## **Disparity tuning of cyclopean visual mechanisms.**

Scott B. Stevenson, Clifton M. Schor and Lawrence K. Cormack

School of Optometry, University of California at Berkeley, Berkeley CA 94720  
and Department of Psychology, University of Texas at Austin, Austin TX 78712

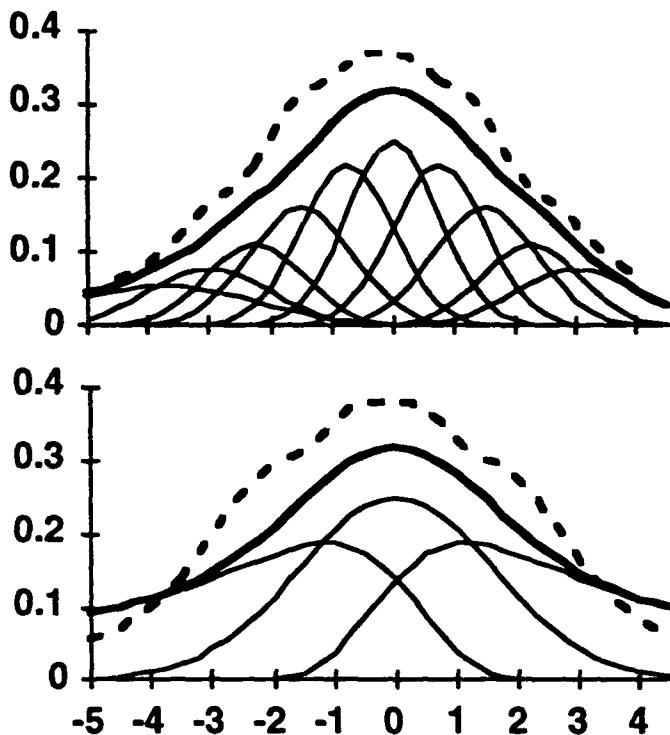
### **ABSTRACT**

Tuned mechanisms or "channels" have been demonstrated in many aspects of human vision, and their characteristics span a continuum from a small set of broadly tuned channels (as in the spectral tuning of cone mechanisms) to a large array of narrow channels (as in the spatial tuning of cone mechanisms). The optimal number and tuning widths of channels for a given dimension depends on a trade-off between an economy of processor resources and the avoidance of metamerism. A small number of broad channels requires a small investment in processor resources and can support fine discriminations but is subject to metameric confusions. A large number of narrow channels requires a greater investment in processor resources but allows for the representation of multiple values on the tuning dimension (e.g. transparency). In the context of stereopsis and vergence control, single unit recordings have provided evidence that disparity tuned mechanisms cover the range from closely spaced, narrow channels ("tuned" cells) to widely-spaced, broad channels ("near/far" cells). In principal, near/far mechanisms should be sufficient to control vergence and allow for fine stereoacuity right around the horopter. Tuned mechanisms might be required for fine disparity discriminations off the horopter and for the perception of stereo transparency. We have investigated the disparity tuning characteristics of binocular visual mechanisms which mediate 1) the psychophysical detection of surfaces in dynamic noise stimuli and 2) the involuntary oculomotor vergence responses to such surfaces. We have found evidence that both perceptual and oculomotor systems involve a large set of narrowly tuned mechanisms with inhibition between neighboring channels. A model is developed which clarifies the non-obvious relationship between measured tuning functions and characteristics of underlying channels.

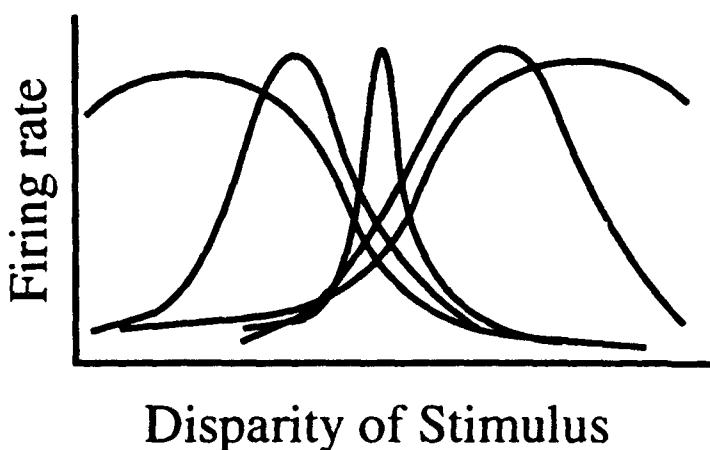
### **1. INTRODUCTION**

The general notion of sensory channels is that there is a set of mechanisms which are individually tuned on some dimension- color, spatial frequency, disparity and so on<sup>1</sup>. Their summed responses produce an envelope of sensitivity across the dimension. Their differential responses allow fine discriminations along the dimension, even though the channels themselves may be broad. Figure 1 shows an example of two sets of filters, with their envelope of sensitivity and their individual tuning functions.

The filter sets shown in Figure 1 represent two extremes- one set has many narrow channels while the other has just three broad channels. The filters have been chosen to produce nearly identical envelopes and similar discrimination functions (differing only by a scale factor). This is intended to make two points. First, given empirically determined envelope and discrimination functions, one can make only limited conclusions about the underlying filter set because there are a large number of possible configurations which give rise to the same general characteristics. Second, the principal difference between these two extremes in filter design may be their susceptibility to metamerism. In the context of stereopsis, this would mean that with many narrow channels one could see multiple, transparent surfaces without confusion, whereas with broad channels the depth information from these multiple surfaces would be averaged together. There may also be a cost to having many narrow channels, however, if one considers that many sensory dimensions are multiplexed in the visual system



**FIGURE 1:** Two example filter sets with similarly shaped envelope sensitivities (bold lines) and discrimination functions (dashed lines) but very different underlying channel structures (thin lines). Envelope and discrimination functions were computed with a quadratic summation rule. Discrimination functions have been scaled separately for the two examples.



**Figure 2:** Schematic of neural responses to disparate stimuli, adapted from Poggio et al. (1988). Single unit responses range from tightly tuned to broadly opponent for disparity.

and are sharing a finite neural resource. More channels for disparity might mean fewer for orientation or spatial frequency, for example. For the construction of artificial vision systems, this form of resource allocation is an important design consideration in which the filter sets are chosen to match the demands of the most common tasks. For the study of natural vision systems, discovering the ways in which neural resources are allocated sheds light on the kinds of tasks which the system as a whole is best suited to.

Our interest has been in comparing disparity tuning for stereopsis and for disparity-driven vergence movements. For stereopsis, there is a clear advantage in having many narrow channels, since the world presents targets at many depths all at once, often with one form or another of transparency. For example, when one looks through foliage at a more distant scene, there are multiple depths represented within each small patch of the visual field. A coarse representation of disparity would lump them all together, whereas a fine representation allows one to see all the depths distinctly and perhaps better focus attention on the objects of interest.

Disparity vergence may have somewhat different requirements for disparity channels. Since vergence can have only one value at a time, all of the disparity information in the visual field must in some way get reduced to a single vergence demand. It might be that a coarse representation could handle this job adequately without a large investment in neural resources<sup>2</sup>. On the other hand, relying on a coarse representation could lead to errors under conditions of transparency or otherwise complex depth distributions, such that vergence fixation would end up at the average target position instead of at a particular target's location. A large number of channels could allow for more selective vergence control.

The physiology of disparity processing provides some evidence for both narrowly tuned mechanisms and broad mechanisms<sup>3,4</sup>. Figure 2 is a schematic which is intended to capture the disparity tuning profiles reported by Gian Poggio and others for cells in areas

V1 and V2. The curves range from the broad, asymmetric Near/Far cells to the narrow, symmetrical Tuned cells. The current thinking seems to be that there is a continuum of cell types between these extremes. It is still a mystery how these various cell types might contribute to the very distinct processes of stereopsis and disparity vergence. It is plausible, however, that stereopsis and vergence might rely on different subsets of these mechanisms and thus show different tuning characteristics. Our results suggest that in fact they have similar disparity tuning characteristics, at least when the stimuli are dynamic random element stereograms.

## 2. PSYCHOPHYSICAL STUDIES OF DISPARITY CHANNEL CHARACTERISTICS:

Figure 3 shows a cartoon of the stimulus configuration used in both psychophysical and oculomotor experiments. A circular field of dynamic random dots was viewed haploscopically in a darkened room. The observer fixated a stationary target in the middle of the field and the disparity of the random dots was varied by shifting one eye's image laterally behind the aperture. The contrast of each frame was 80% and the frame rate was 60Hz. In cases where a transparent stimulus was used, the two disparity values were presented on alternate frames so that each surface in depth was presented at 30 Hz. This is fast enough that observers don't see any motion in depth between the surfaces. The interocular correlation of each surface was controlled by varying the proportion of matching dots in the left and right images on a frame by frame basis. This has the effect of altering the signal strength of a cyclopean surface and is analogous to luminance contrast for spatial vision. The functions drawn below the surfaces represent the correlation signals presented to observers in this situation.

Figure 4 is an autostereogram which can be viewed by free fusing the eyeball icons in the center. The stereogram portrays a vertical square wave disparity grating. At the top of the figure the dots match perfectly and so the interocular correlation is 100%. It decreases smoothly down to 0 at the bottom, where the matches are completely random and no surface is perceived. The reader might be able to get a sense of where his/her own correlation detection threshold is by noticing how far down the figure he/she can still see the grating. The first thing we measured in our psychophysical experiments was the

correlation threshold for detecting a flat surface against a null target of 0 correlation. The correlation threshold depends on several factors, in particular on the disparity of the surface and on the overall number of dots presented across space and time<sup>5-7</sup>. Most of our observers have thresholds of around 5-10%, which is in the vicinity of the lower pair of eyeball icons in figure 4.

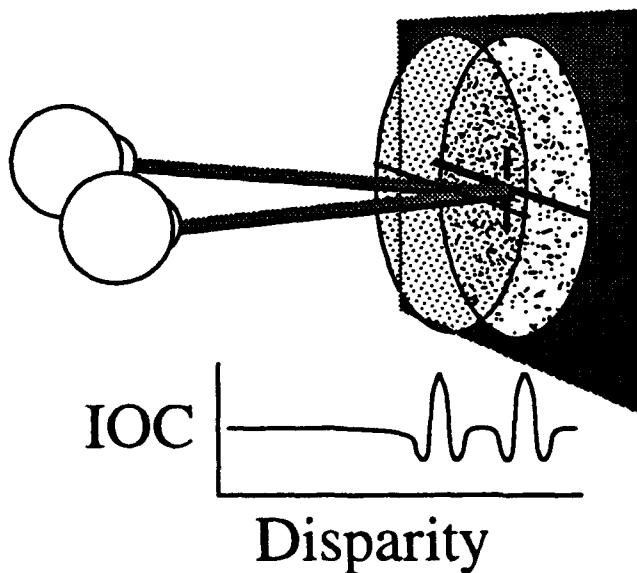


Figure 3: Schematic of dynamic random dot stimulus showing two surfaces in a transparency configuration. Inset: Schematic cross-correlation function of stimulus.

Figure 5 shows baseline correlation threshold data for three observers as a function of disparity. Correlation is plotted on a log scale. The endpoints of each observer's data represents the Dmax for these conditions and the function as a whole indicates the overall envelope of sensitivity to correlation across disparity. In order to get some idea of the channels which produce this envelope, we adapted observers to a fully correlated surface at a particular disparity and measured the baseline again in the adapted state. Figure 6 shows the effect on baseline threshold of adaptation to a surface at zero disparity for one observer. Notice that the threshold is elevated at the adapted region and that there is a marked

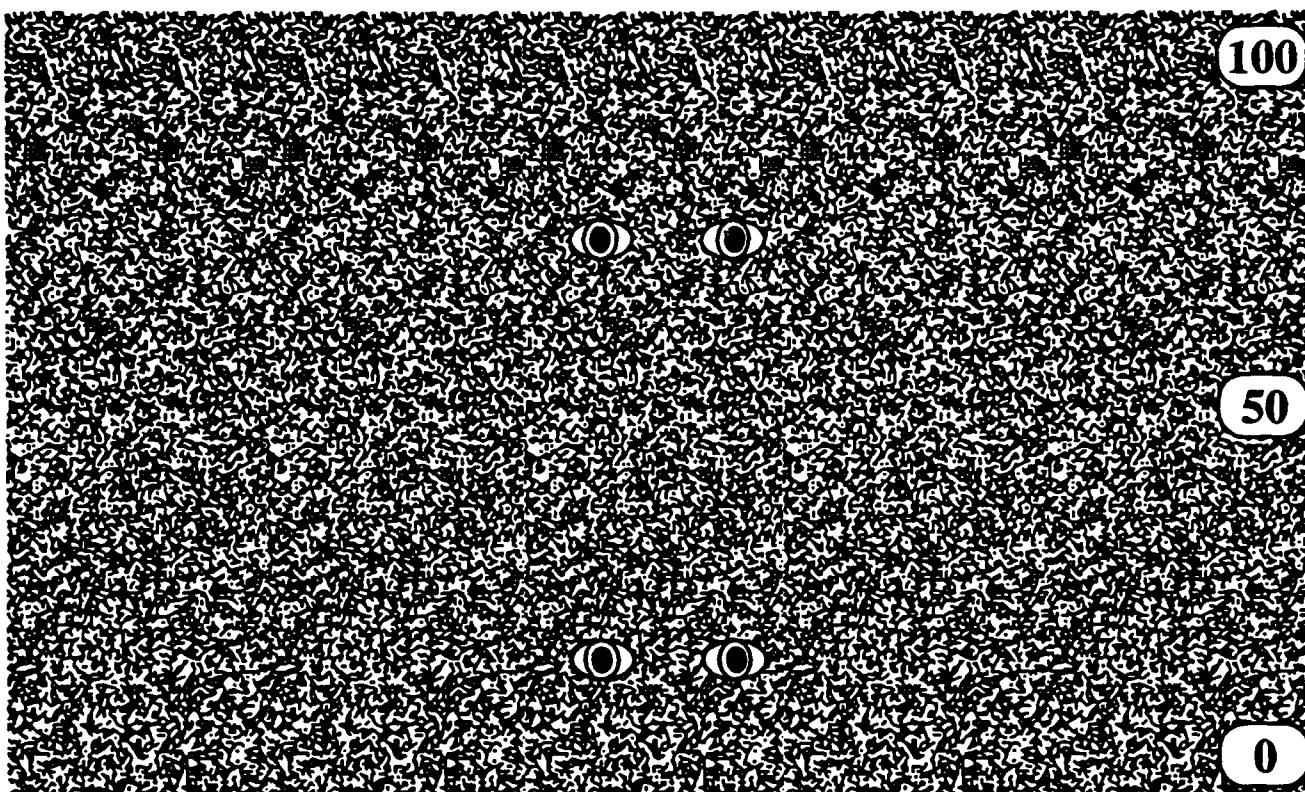


Figure 4: Autostereogram illustrating the effect of interocular correlation on the appearance of a random-dot stereogram. Correlation is 100% at the top and ramps down to 0% at the bottom. By free fusing the small "eyes" near the top of the figure, the observer should see a square wave disparity grating (series of vertical strips standing out in relief). The grating becomes increasingly difficult to see nearer the bottom of the figure and the pattern as a whole becomes fuzzy.

facilitation in the surrounding region. From this comparison we generated tuning functions by taking the difference in log correlation thresholds before and after adaptation. Figure 7 shows some examples of the disparity tuning functions obtained.

The upper panel of Figure 7 shows tuning functions for adaptation at zero disparity for three observers, and the lower panel shows functions for adaptation at 10 arcmin Near disparity for the same three observers. In every case the functions peak at the disparity of adaptation and show some facilitation for surrounding disparities. Tuning functions derived in this way are not pictures of the channel profiles and they certainly are not pictures of single unit responses, though in some cases they may resemble them. In this case, the tuning functions we obtained look remarkably like the single unit functions reported by Poggio and colleagues. However, in order to interpret data such as these in terms of underlying channels or filters, it is helpful to construct a channel model and compare the adaptation tuning functions produced by the model to those produced by human observers.

Figure 8 shows the channels used to model the adaptation data of Figure 7. One channel is plotted more darkly than the rest to show the profile used. Each channel is a difference of two Gaussians, with a one to six ratio in amplitude and a six to one ratio in width so that the overall function is balanced. The channel sensitivity falls off with disparity away from the horopter and channel width increases to compensate, so that the overall sensitivity of each channel is the same. In this figure, the channel sensitivities are shown after adaptation to a stimulus at 15 arc min. Channels that peak near 15 min are reduced in sensitivity and those with negative sidelobes near 15 min are potentiated by the adaptation.

The overall sensitivity of this set of filters or channels was determined by combining their outputs with quadratic combination rule, resulting in a baseline response comparable to the human data. The model nicely captures the envelope of sensitivity, but there are any number of channel configurations which might give the same result. The more critical comparison is of the human and model disparity tuning profiles.

Figure 9 shows three tuning functions produced by the channels shown in Figure 8. Each tuning function peaks at the locus of adaptation and has facilitatory surround, as did the human data. The tuning functions get broader with increasing disparity and they also become asymmetric, as did the human data. This asymmetry arises from the falloff in channel sensitivity with disparity, not from asymmetric channel profiles. There were two main conclusions reached from this modeling: first, that the channels had to be of the "Tuned" variety because asymmetric channels always failed to produce the return to baseline beyond the adaptation site; and second, that the channels had to be narrow and numerous because the peak of the disparity tuning function was always at the site of adaptation and the tuning functions were uniformly narrow. See Tyler in this volume for a discussion of the relationship between measured tuning functions and underlying channels<sup>8</sup>

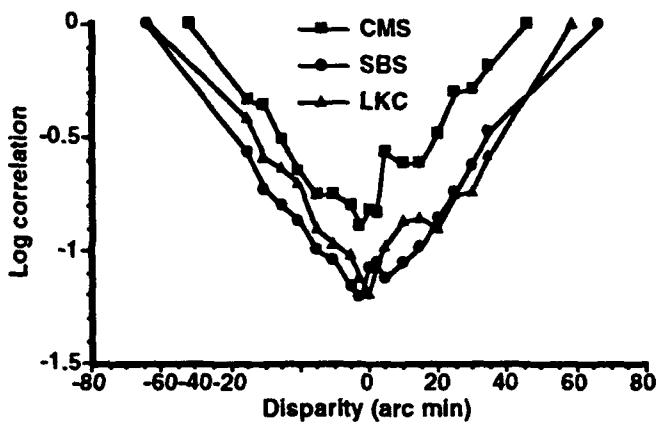


Figure 5: Unadapted correlation thresholds

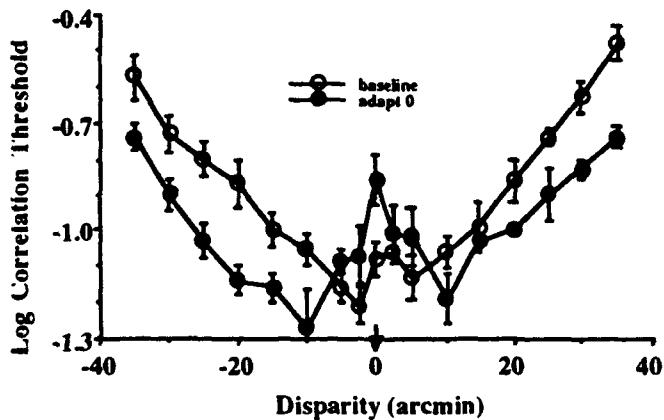


Figure 6: Effect of adaptation at zero disparity on correlation thresholds.

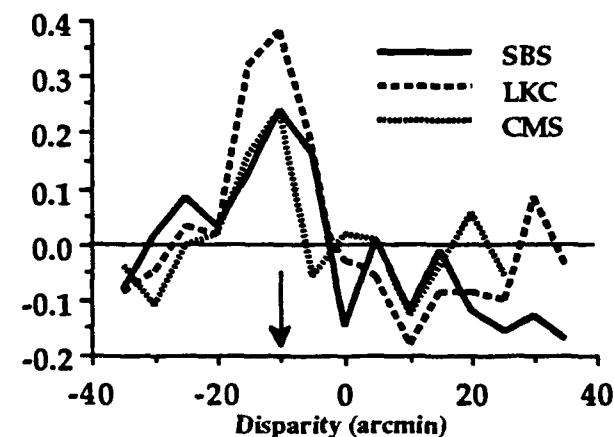
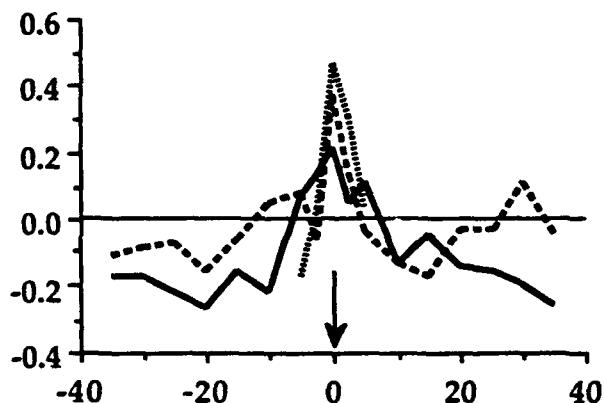


Figure 7: Tuning functions derived from adaptation to a particular disparity. Data are shown for three subjects. Top: threshold elevation (or reduction) after adaptation to zero disparity. Bottom: Same, after adaptation to -10 arcmin.

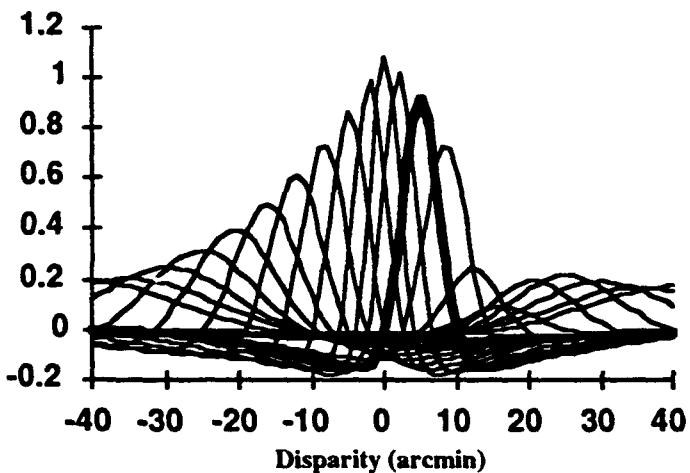


Figure 8: Disparity tuned channels used in modeling.

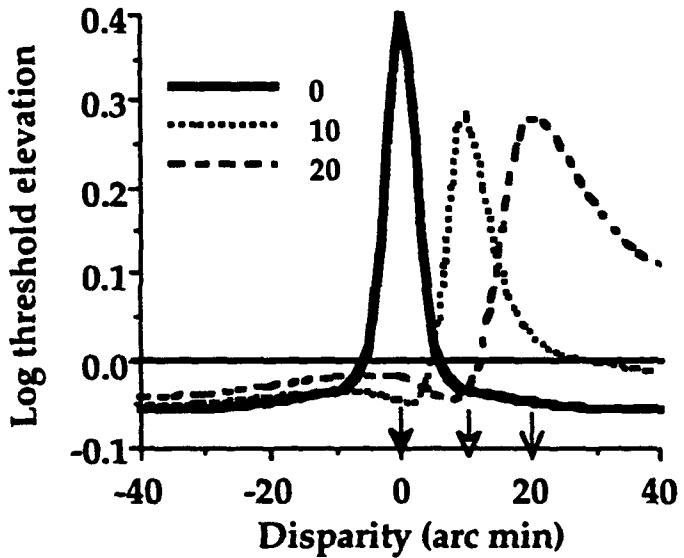


Figure 9: Tuning functions from a multi-channel model developed to account for data from the adaptation experiment. Arrows on horizontal axis indicate adaptation loci.

vertical scaling was changed to match up the two functions. The responses to transparent stimuli show clearly both the narrow tuning and the surround inhibition in the disparity tuning profiles.

### 3. OCULOMOTOR STUDIES OF DISPARITY TUNING EFFECTS IN VERGENCE RESPONSE:

A variant of the threshold summation procedure was used to examine the disparity tuning of disparity vergence responses. We measured observers' vergence responses to transparent surfaces and compared them to responses to single surfaces to see how two simultaneously presented surfaces would interact. Of particular interest was responses to the condition in which two surfaces are presented on the

In a separate series of experiments we used a threshold summation paradigm<sup>9</sup> to look at disparity tuning functions with the same stimuli. In this case, we measured correlation thresholds for detecting a pair of surfaces presented in the transparency configuration and we varied the disparity and relative signal strength of each surface to assess their interaction at threshold. Figure 10 represents the paradigm in a schematic form. On the left side are threshold contours corresponding to three disparity separations. The straight diagonal line is obtained when two stimuli simply add together, as for example when they have the same disparity. When the two components of the stimulus are detected independently, the nearly square function is obtained. This occurs when a component falls on a cross-over point or falls completely outside of the channel sensitivity. When the two components are detected by opponent mechanisms, the highly bowed threshold contour is obtained. This occurs when each component falls in the negative sidelobe of the channel that is detecting the other component. These cases are schematized on the right hand side of the figure, where the gray arrow represents one component of the transparent stereogram and the three black arrows represent the three cases just described. The tuning function is derived by first measuring the degree of bowing in the threshold contour for each disparity separation, and then comparing that to the case of independence.

The results from this threshold summation experiment agree quite well with the adaptation experiment, as shown in Figure 11. Tuning functions obtained in the adaptation and threshold summation paradigms are compared for one observer at 10 arc min near and for another observer at zero disparity. Only the

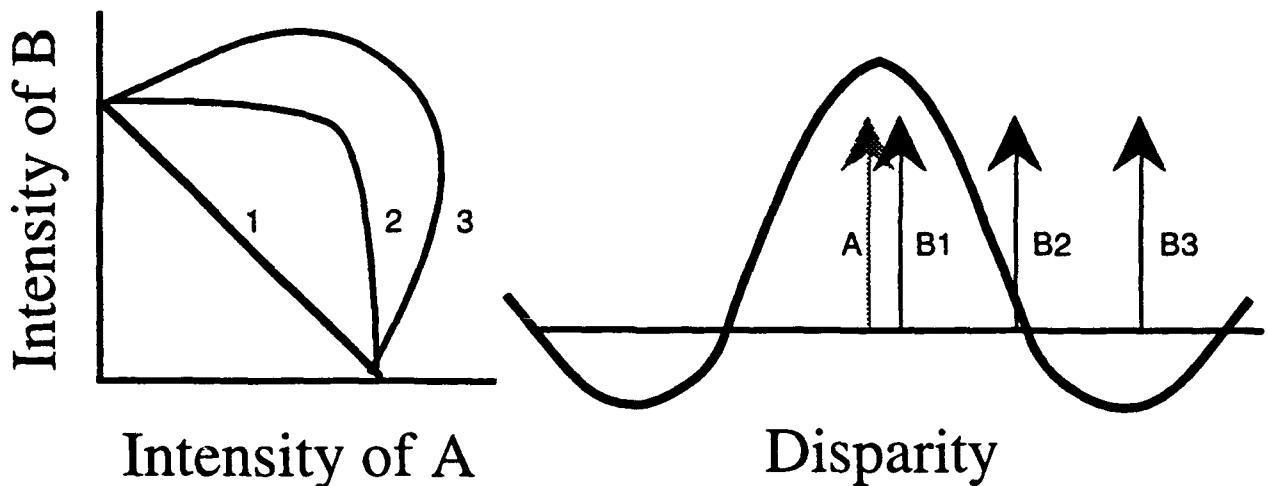


Figure 10: Derivation of tuning functions from threshold summation paradigm. Each numbered curve in the left panel represents thresholds for simultaneous detection of stimulus A and the correspondingly numbered stimulus B in the right panel. Shape of the threshold contour relative to that expected for probability summation reveals the nature of the interaction between stimuli.

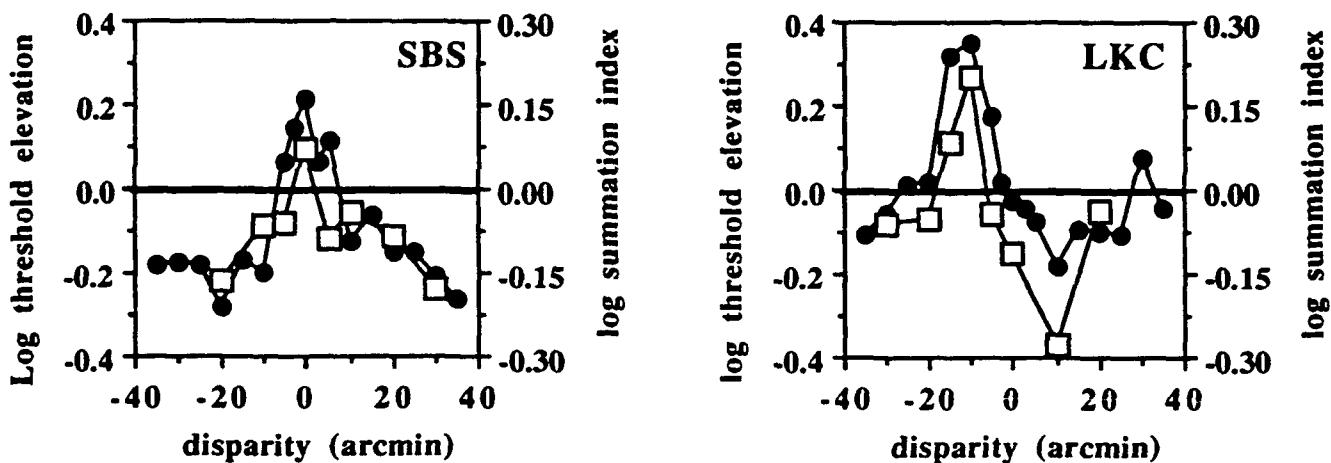
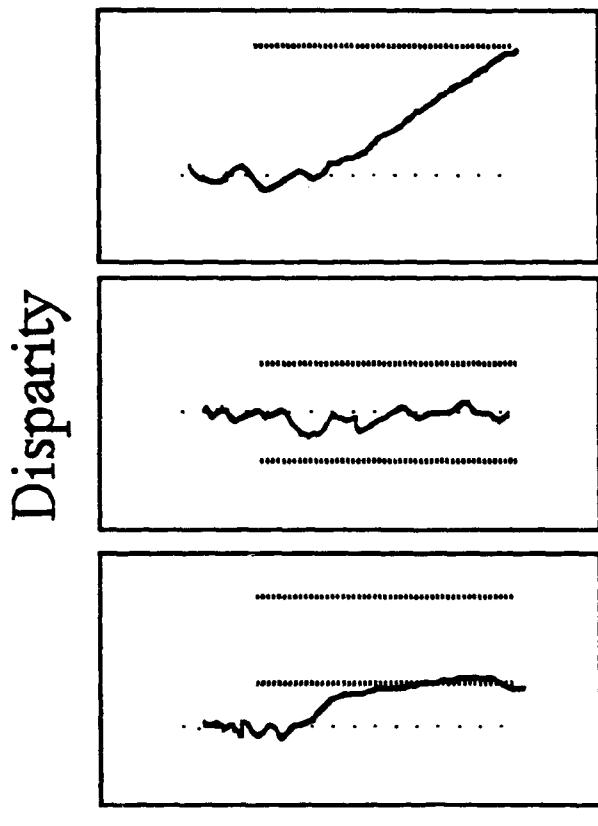


Figure 11: Comparison of tuning functions from adaptation (circles) and from two plane summation (squares) studies. Left, subject SBS adapted at 0 disparity. Right, subject LKC adapted at -10 arcmin.

same side of the horopter. We reasoned that if the channels measured psychophysically were also controlling disparity vergence, then the inhibitory surrounds we observed should cause a reduction in vergence response in this case. If the channels driving vergence were like the near/far type cells, with broad sensitivity for all disparities on one side of the horopter, then summation should be observed. In either case, we expect opponency when surfaces are presented simultaneously on each side of the horopter.

The observers' eye movements were tracked with an SRI dual-Purkinje eye tracker while they watched a dynamic random element display that was very similar to the one used in the psychophysical studies. Observers were instructed to hold their vergence steady when the stimulus came on, but the fixation mark was weakened considerably so that measurable involuntary responses were produced. Observers had nonius lines to monitor their own eye position and pressed a button to initiate a trial when the nonius marks were aligned.



## Time

Figure 12: Schematic of time course for stimulus presentation and vergence response in the two plane step vergence paradigm. Top: a single surface presented at some disparity. Middle: two surfaces which straddle the horopter. Bottom: two surfaces on the same side of the horopter.

velocity between transparency and single surface conditions. The sign of the difference is adjusted so that the tuning function indicates the effect which the zero disparity surface has on response to the other surface. This function is shown as the dashed line in the right panels (B, D). When plotted this way, the tuning function looks very much like those obtained psychophysically for zero disparity and the conclusions are much the same: mechanisms sensitive to stimuli at zero disparity have a strong inhibitory effect on surrounding mechanisms.

Now consider the function which represents the case where one surface is fixed at 10 arcmin (filled squares). When the variable surface is presented on the opposite side of the horopter, the response is almost zero, as expected under any reasonable model. However, when the variable surface is on the same side but with larger disparity, the response is also less than expected. Since the disparity of one surface is getting bigger, we would expect that the velocity should continue to rise but instead it stays flat. This interaction is represented in the right panels by the tuning curves marked 10 and -10. This shows that under conditions of transparency, the vergence response is lower than for a single surface, even when both components are on the same side of the horopter. This suggests to us that tuned, probably mutually inhibitory mechanisms are responsible for the vergence responses as well as the

The three panels in Figure 12 show examples of various stimulus configurations presented to the observers. In some cases, there was a single surface presented, in others there were two surfaces, and these might be on the same or opposite sides of the horopter. The response traces shown here are schematic, not real vergence traces, but they are not too different from what we recorded. The response was quantified by measuring the initial velocity of the vergence: that is, the slope of the response right after it begins.

Figure 13 summarizes the results for one observer. The left panels (A,C) show vergence velocities to stimuli with either one surface (the heavy line) or with two surfaces (the other four curves). The right panels (B, D) show tuning functions that were derived from the velocity data. For clarity, responses to near and far disparities have been plotted separately in the top and bottom panels, respectively. For the transparency conditions, each curve shows responses when one surface is fixed at some disparity (shown in legend) and the other is presented at various other disparities (shown on horizontal axis).

Notice that the response to a single surface (heavy line with no symbols) increases as disparity increases. The correlation was always 100%, so in that sense the signal strength is constant, but disparity also contributes to signal strength in the vergence system. Now consider the function obtained for the condition in which one surface was always at zero disparity (diamonds). The presence of this zero disparity surface virtually killed the response to other surfaces, even though by themselves they would have been effective stimuli. This was translated into a tuning function by taking the difference in initial vergence

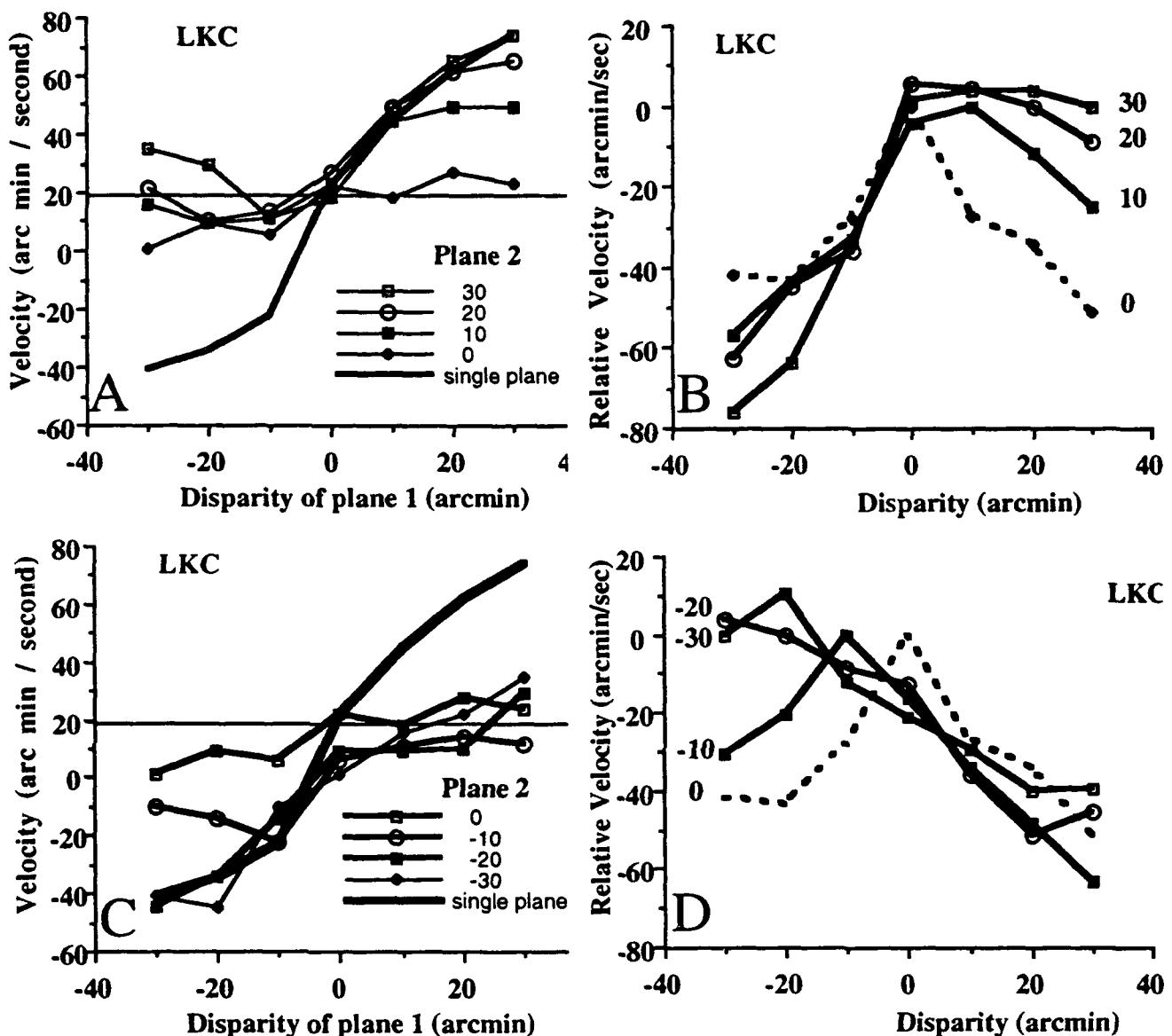


Figure 13: Derivation of tuning functions from vergence responses to one and two planes of random dots. For clarity, functions for far (A,B) and near (C,D) disparity have been plotted separately. A,C vergence velocities to two planes whose disparities are shown on horizontal axis and in legend. B,D, relative velocities computed from difference of two-plane and one-plane curves in left panels.

psychophysical responses to these stimuli. Results for 20 and 30 arc min are less clear, but we would have to test much larger disparities in order to see if the response attenuates in the same way or continues to rise as would be expected for a near/far system.

#### 4. CONCLUSIONS

Based on these psychophysical and oculomotor results, we believe that the disparity vergence system gets signals from the same or a very similar set of correlation sensitive filters that serve as a precursor to stereopsis. These filters are characterized by their relatively narrow tuning and by their inhibitory

surrounds in the disparity domain. While we have not fully developed a channels model for vergence control, as we did for psychophysical responses, the similarity in the tuning functions obtained under the two paradigms suggests a similar set of underlying filters. It was argued earlier that the principle advantage of such a filter set may be the avoidance of metamerism confusion. With respect to involuntary vergence control, it appears that the system is designed to hold fixation on local maxima in the correlation vs. disparity function rather than to keep fixation at some more global average position. Most surfaces in the real world are opaque and present a single, clear signal to hold fixation. In cases of transparency (e.g., objects viewed below the surface of water or through a veil of foliage), the involuntary fixation control system tends to keep fixation at one depth plane or another instead of in empty space at some weighted mean of the disparity inputs. Such a "winner-take-all" strategy is only useful if there is a sufficiently large set of filters that the system gets independent signals from many depth planes, so as to allow selection of a particular target for refixation. Once fixation changes, the same strategy allows for maintained fixation despite the presence of salient targets at other depths. This may be another reflection of the general tendency of the human visual system to have resources focused into a particular area (e.g. the enhanced spatial sensitivity of foveal vision, the enhanced stereoacluity and correlation sensitivity near the horopter) rather than distributed generally. Whether such a design for vergence control is appropriate for machine vision or not depends on the design of the sensing components and on the degree to which the task involves focused attention and an episodic, 'clumpy' environment.

## 5. ACKNOWLEDGMENTS

This work was supported by grant number EY08882 from the National Eye Institute. Portions of this work have been published previously in *Vision Research*<sup>10</sup> and *Visual Neuroscience*<sup>11</sup>.

## 6. REFERENCES

1. Regan, D. "Visual information channeling in normal and disordered vision." *Psychological Review*, 89, 407-444. 1982.
2. Jones, R. "Fusional vergence: sustained and transient components." *Am. J. Optom. Physiol. Opt.* 57, 640-644. 1980.
3. Poggio, G. F. , Gonzalez, F. and Krause, F. "Stereoscopic mechanisms in monkey visual cortex: binocular correlation and disparity selectivity." *Journal of Neuroscience* 8, 4531-4550. 1988.
4. LeVay, S. and Voigt, T. "Ocular dominance and disparity coding in cat visual cortex." *Visual Neuroscience* 1, 395-414. 1988.
5. Tyler C. W. "Binocular cross-correlation in time and space." *Vision Research* 18, 101-105. 1978.
6. Cormack, L. K., Stevenson, S. B. and Schor, C. M. "Spatial and temporal integration of cyclopean vision." *Invest. Ophthalm. Vis. Sci. Suppl.* 33, p.1368. 1992
7. Cormack, L. K. The Binocular Processing of Random Element Stereograms of Variable Interocular Correlation. Ph. D. Thesis, University of California at Berkeley. 1992.
8. Tyler, C. W. "Computational reconstruction of the mechanisms of human stereopsis." This volume. 1993
9. Kranda, K. and King-Smith, P. E. "Detection of coloured stimuli by independent linear systems." *Vision Res.* 19, 733-745. 1979.
10. Stevenson, S. B. , Cormack, L. K. , Schor, C. M. and Tyler, C. W. "Disparity tuning in mechanisms of human stereopsis." *Vision Res.* 32, 1685-1694. 1992.
11. Cormack, L. K., Stevenson, S. B. and Schor, C. M. "Disparity-tuned channels of the human visual system." *Visual Neuroscience* 10, 585-596. 1993.

## **Standard model of color vision: problems and an alternative**

**Russell L. De Valois**

**Psychology Department and Vision Science Group  
University of California, Berkeley**

### **Abstract**

The "Standard Model" of early color processing postulates: an achromatic magno LGN non-opponent pathway summing the outputs of the L and M cones; a red-green parvo LGN opponent cell system differencing L and M cones; and a yellow-blue or tritan parvo LGN opponent cell system differencing S from the (L+M) cones. A number of psychophysical and perceptual findings, however, do not agree with this Standard Model, and we have suggested an alternative.

Our model diverges from that above in three fundamental ways: 1. that L-M and M-L cells do not constitute the "red-green" system, but serve as the principal inputs to both the red-green and yellow-blue systems; 2. that S-LM and LM-S opponent cells do not constitute the yellow-blue system, but rather combine at a third stage with the LM opponent cells in different ways to produce both the red-green and the yellow-blue systems, serving a modulatory role to break the one effective LGN response axis into separate red-green and yellow-blue perceptual color axes at some cortical site; 3. in addition to chromatic information, the parvo opponent cells (as well as the magno cells) carry intensity information, the chromatic and intensity components being separated at the third stage.

### **Characteristics of the standard model.**

For years, starting in the last century, there were fierce disputes in the literature over conflicting theories of color vision, primarily between the followers of Helmholtz and those of Hering. Over the last few decades, however, widespread agreement has developed on a general model of color vision, which we refer to as the Standard Model. It in effect combines the essential features of both Helmholtz's and Hering's ideas, although there is still much disagreement about particulars.

The Standard Model assumes three cone types, containing three different photopigments of broad spectral sensitivity. This stage is assumed to determine the fundamental three-dimensionality of normal human color vision. However, there is not, as in a strict Helmholtzian model, a direct path from each of these cone types to central regions. Rather, the cone outputs are combined in three different ways to form three paths or mechanisms. One path is that of the spectrally non-opponent cells which sum the outputs of the L and M cones and carry luminance information through the magnocellular geniculate layers to the cortex. A second path, forming part of the parvocellular geniculate layers, consists of the red-green opponent cells, which difference the outputs of the L and M cones: L-M and M-L. The third path consists of the yellow-blue parvo cells, which difference the outputs of the S cones from that of the L and M cones combined: S-LM and ML-S. The basic opponent character of our color vision (emphasized by Hering) is, by this model, due to the opponent nature of these spectrally-opponent cells.

### **Problems with the standard model.**

1. The fundamental three-dimensionality of (normal) human color vision -- trichromacy -- has been recognized from the earliest scientific investigations of human vision. For almost as long, from the time of Thomas Young in 1802, it has been assumed that this striking limitation to our visual capabilities is due to the presence of only three cone types, with three different photopigments. It has long been known that we in fact have a fourth type of receptor, rods, but it has been implicitly assumed that the rod signals ascend through a separate path way to the brain, one that has nothing to do with color vision. Recent evidence indicates that both of these assumptions -- that color-normals have but three cone pigments, and that there is a separate rod path to the brain -- are false. As a consequence, we need to reassess the cause of trichromacy, to examine again the possible site of the three-dimensional limit to our color vision.

For a number of years, psychophysical evidence has been accumulating that there exist several middle-to-long wavelength (ML) pigments among individuals with normal color vision, not just the two with peaks at 530 and 561 nm that are assumed by the Standard Model. There appear to be a number of photopigments with absorption peaks between 530 and 561 nm, spaced at roughly 5 nm intervals (Neitz, Neitz & Jacobs<sup>1</sup>). Someone with 530 and 556 nm pigments would have slightly different color vision than one with 530 and 561 nm pigments, but both would still be within the normal range. Given the presence of genes for several different ML pigments, one would expect that a certain percentage of females might inherit, on their two X-chromosomes, two different L-cone pigments or two different M-cone pigments. They would then have not 3, but 4 or even 5 cone pigments. The multimodal anomaloscope settings made by females in a large sample (Neitz & Jacobs<sup>2</sup>) indicate that this may indeed occur. However, females who clearly have more than three cone pigments nonetheless make trichromatic color matches (Nagy, MacLeod, Heyneman & Eisner<sup>3</sup>). It is also clear that rods and cones are both functional under most daylight visual conditions, that they are both present everywhere in the retina except for a tiny region in the foveola, and that rods do not have a separate path to the brain, but rather feed into the same ganglion cells that pick up from cones. But again, studies show that we still have only a three-dimensional, trichromatic, visual system even with stimuli which clearly activate both rods and cones.

2. Our early recordings from single units in the monkey lateral geniculate nucleus (De Valois; De Valois, Abramov & Jacobs<sup>4</sup>) found evidence for spectrally opponent cells of two different varieties that, to a first approximation, appeared similar to the red-green and yellow-blue perceptual opponent channels of Hering and of Hurvich and Jameson<sup>5</sup>. However, as we pointed out at the time, and as the more recent recordings of Derrington, Krauskopf & Lennie<sup>6</sup> make even clearer, there is not complete agreement between the LGN opponent-cell chromatic axes of optimal response and the perceptual opponent axes. Thus the S-LM opponent cells respond preferentially along the tritan axis, which does not precisely coincide with the perceptual yellow-blue axis; and the L-M and M-L opponent cells respond preferentially along an orangish-red to cyan axis when modulated around white, rather than along a true perceptual red-green axis. Thus while there is an approximate agreement between geniculate opponent cells and perception, as incorporated in the Standard Model, a clear second-order discrepancy exists.

3. The Standard Model postulates that the yellow-blue color system is based on differencing the outputs of the S cones from the combination of the LM cones, the S-LM opponent cells. Several current anatomical techniques allow one to identify the S cones histologically (Ahnelt, Kolb & Phlug; Curcio, Allen, Sloan, Lerea, Hurley, Klock & Milam<sup>7</sup>). Examination of both human and macaque retinae with these techniques clearly confirms what had long been believed from psychophysical tests, namely that there are very few S cones in the retina. Overall, they form about 6% of the total and are even less prevalent in the fovea, being virtually if not totally absent from the central foveola. It seems intrinsically implausible that half of our color vision, the yellow-blue system, is based on the outputs of just 6% of the cones, and the other half, the red-green system, on the other 94% of the cones. On the contrary, the red-green and yellow-blue systems seem perceptually balanced relative to each other.

4. If the blue-yellow perceptual system were based on the outputs of the S cones (often referred to as "blue" cones!), one would expect that their elimination would lead to a loss of the percept of blue (and yellow). The long-wavelength half of the spectrum should look red and the short-wavelength half, green. However, a number of psychophysical experiments show that that is not the case. Various procedures, such as confining the stimulus to the S-cone-free central foveola, or presenting very small, brief flashes of light, can minimize the contribution of S cones. The effect of all of these is just the opposite to that predicted by the Standard Model, namely, the short wavelengths half of the spectrum under these circumstances appears not green but blue (Boynton, Schafer & Neun; Drum<sup>8</sup>). The same conclusion is reached in studies of rare unilateral tritanopes, individuals who are totally lacking S-cone function in one but not in the other eye (Ohba & Tanino; Alpern, Kitahara & Krantz<sup>9</sup>). It is not very meaningful to ask an ordinary color-blind individual what colors he sees, but these unilateral tritanopes can directly compare the color of monochromatic lights in their normal and tritanopic eyes. What these observers report is that all short wavelengths appear blue in the eye lacking S cones, which is again the opposite of that predicted by the Standard Model.

5. In the Standard Model, intensity information is carried just by spectrally non-opponent cells. However, spectrally-opponent cells respond to intensity as well as to color changes, effectively multiplexing chromatic and intensity information. Since these parvocellular opponent cells constitute about 80-90% of the path from retina to cortex, it seems unlikely in the extreme that the intensity information they carry is not utilized by the visual system.

6. A set of findings perhaps related to point 4 above shows a distinct asymmetry **within** the red-green system and **within** the yellow-blue system, a situation not at all to be expected from the Standard Model. Large spots of light presented in the periphery appear very similar to the same wavelengths in the fovea, but small peripheral spots appear quite different. Abramov, Gordon and Chan<sup>10</sup> showed that as spot size is decreased, lights that had appeared yellow or green become desaturated. All short and middle wavelength lights now appear blue, and longer wavelengths, red. It would thus appear that while red and green are usually in opponent relation to each other, as Hering emphasized, they are not necessarily just mirror-image components of a single system. The same is true for blue and yellow.

### **Our multi-stage color model.**

In an attempt to deal with some of the problems with the Standard Model (especially #2-6 outlined above), Karen De Valois and I<sup>11</sup> have developed the broad outlines of an alternative color model, incorporating an additional processing stage. The essential novel feature of our model is that the S-opponent system (S-LM opponent cells) is seen not as constituting the blue-yellow color system but rather as playing a modulatory role in forming both the red-green and the blue-yellow systems. The L-M and M-L opponent cells, in our model, do not form the red-green system, but rather are the primary inputs to both the red-green and the yellow-blue systems.

The first stage of our model consists of 3 or 4 or 5 cone pigments (among individuals with "normal" color vision), but contained in only 3 types of cones: L, M and S cones. That is, although some individuals may have more than one long-wavelength pigment, the neural pathway does not distinguish among the cones containing different L photopigments. Thus trichromacy, we postulate, lies in the presence of only three central neural systems, not in the presence of just three cone pigments. We assume, in line with certain psychophysical evidence, that the overall proportion of the different varieties of cones in the retina is 10L: 5M: 1S.

The second stage of our model has three cone-opponent cell systems, produced by interactions in the retina and feeding up the parvo geniculate path; and one cone-non-opponent system (cells which sum the outputs of L and M cones and feed up the magno path). This is similar to the Standard Model, and it is in agreement with our and others' recording data, except that we do not consider the L-M and M-L cells to be the red-green system, or the S-LM cells to be the blue-yellow system. Rather, these geniculate cells are just seen as an intermediate processing stage. Thus we have at this stage what we may term the  $L_o$  (L-M cone-opponent) cells,  $M_o$  (M-L cone opponent) cells and  $S_o$  (S-LM cone-opponent) cells, in addition to the  $L+M$  and  $-L-M$  non-opponent cells.

The novel feature of our model is that we postulate a 3rd processing stage, at some cortical locus, at which the various cone-opponent types are combined in various ways to produce the perceptual red-green and yellow-blue color systems, and to separate the multiplexed chromatic and intensity information present in the geniculate opponent cells. The postulated interactions at this third stage serve to rotate the cone-opponent axes to make them correspond to the perceptual color axes.

The interactions we postulate to produce perceptual opponent-color space are as follows. As stated above, the  $M_o$  and  $L_o$  cells, constituting 80 to 90% of the visual projection, would form the main inputs to all the color systems. At this third stage, the  $S_o$  cone-opponent cells are added to or subtracted from the  $M_o$  and  $L_o$  systems to rotate the color axes and thus form the perceptual color systems. We postulate that  $M_o + S_o$  cells give blue;  $M_o - S_o$  cells give green;  $L_o - S_o$  cells code yellow; and  $L_o + S_o$  signal red. The response functions resulting from these postulated interactions in fact correspond very well to perceptual color naming of different spectral regions. Note that since  $L_o$  and  $M_o$  are roughly mirror-images of each other (L-M vs M-L), red and green are perceptual opposites, as are yellow and blue.

To illustrate how the modulation of the  $M_o$  and  $L_o$  systems by the  $S_o$  system works, and how it accounts for certain facts that are an embarrassment for the Standard Model, let us consider the blue part of the yellow-blue opponent system. In the Standard Model, this is S-LM (or what we are terming  $S_o$ ); we postulate, on the contrary, that our percept of blue is produced by  $M_o + S_o$ . The S cones have their peak sensitivity at about 440 nm and their sensitivity drops rapidly at longer wavelengths, making them essentially completely insensitive beyond 500 nm. Perceptual unique green is that wavelength that appears to have no blue or yellow in it. By any theory, this must be the point at which the blue-yellow system crosses from excitation to inhibition. But this unique green point is at about 510 to 515 nm, whereas the S-opponent cells peak at 440 nm and cross over into inhibition at about 480 nm. Furthermore, as discussed in points 4 and 6 above, there are circumstances under which the whole spectrum up to even 570 nm is seen as blue -- and those circumstances are ones in which the S cones are absent or non-functional. These facts are quite incompatible with the Standard Model, but they correspond to what is predicted by our model. The  $M_o$  cells, in our model, form the main input to blue, and they by themselves show excitation to all wavelengths up to 560-570 nm. When the S-cone system is non-functional, we would therefore expect that the whole short-wavelength half of the spectrum would appear bluish, as it does. Adding  $S_o$  cells to the  $M_o$  cells has two effects. The excitatory branch of the  $S_o$  cells shifts the blue peak to shorter wavelengths; and the inhibitory branch of the  $S_o$  cells (at 480 nm and above) cancels the excitation from  $M_o$  cells at longer wavelengths, producing a null at about 510-515 nm, thus accounting for the locus of unique green.

The role we postulate for the  $S_o$  cells in the red system also accounts for certain other perceptual facts. Pure red is an extra-spectral color, requiring a combination of long and short wavelengths. Adding  $S_o$  cells to  $L_o$  cells gives the red subsystem the combined long+short wavelength inputs it requires to account for our percept of red, and also for the distinctly reddish appearance of very short wavelengths.

A final feature of our multistage color model is the separation at this third stage of the color and intensity information multiplexed in the responses of opponent cells. The basis for doing this, we postulate, is that parvo geniculate cells respond to both intensity and color variations, but with different receptive fields in the two cases. Thus an  $L_o$  cell fires to light increments and inhibits to decrements, and also fires to red and inhibits to green. An  $M_o$  cell also fires to light increments and inhibits to decrements, but it fires to green and inhibits to red. If two such cells with superimposed receptive fields were added together, ( $L_o + M_o$ ), the chromatic responses would cancel and the intensity responses would sum. On the other hand, a  $-M_o$  cell is inhibited by a light increment and fires to a decrement, and also fires to red and inhibits to green. If such a cell were summed with the  $L_o$  cell, ( $L_o - M_o$ ), the chromatic responses would add and the intensity responses would cancel. Thus by combining the outputs of the various cone-opponent cells in different ways, the visual system could, at this third stage, separate chromatic and intensity information. We postulate that the non-opponent (magno) cells carry only intensity information, and in fact account for the photopic luminosity function as measured by flicker photometry. But intensity information is also carried by the spectrally-opponent cells, and this accounts for the added brightness of long- and short-wavelength lights.

## References

1. Neitz, M., Neitz, J. & Jacobs, G.H. (1991) Spectral tuning of pigments underlying red-green color vision. *Science* 252: 971-974.
2. Neitz, J. & Jacobs, G.H. (1986) Polymorphism of the long-wavelength cone in normal human colour vision. *Nature* 323: 623-625.
3. Nagy, A.L., MacLeod, D.I.A., Heyneman, N.E. & Eisner, A. (1981) Four cone pigments in women heterozygous for color deficiency. *J. Opt. Soc. Am.* 71: 719-722.
4. De Valois, R.L. (1965) Analysis and coding of color vision in the primate visual system. *Cold Spring Harbor Sym. Quant. Biol.* 30: 567-579; De Valois, R.L., Abramov, I. & Jacobs, G.H. (1966) Analysis of response patterns of LGN cells. *J. Opt. Soc. Am.* 56: 966-977.
5. Hurvich, L.M. & Jameson, D. (1956) Some quantitative aspects of an opponent-colors theory - IV. A psychological color specification system. *J. Opt. Soc. Am.* 46: 416-421.
6. Derrington, A.M., Krauskopf, J. & Lennie, P. (1984) Chromatic mechanisms in lateral geniculate nucleus of macaque. *J. Physiol. (Lond.)* 357: 241-265.
7. Ahnelt, P.K., Kolb, H. & Pflug, R. (1987) Identification of a subtype of cone photoreceptor, likely to be blue-sensitive, in the human retina. *J. Comp. Neurol.* 255: 18-34; Curcio, C.A., Allen, K.A., Sloan, K.R., Lerea, C.L., Hurley, J.B., Klock, I.B. & Milam, A.H. (1991) Distribution and morphology of human cone photoreceptors stained with anti-blue opsin. *J. Comp. Neurol.* 312: 610-624.
8. Boynton, R.M., Schafer, W. & Neun, M.A. (1964) Hue-wavelength relation measured by color-naming method for three retinal locations. *Science* 146: 666-668; Drum, B. (1989) Hue signals from short- and middle-wavelength-sensitive cones. *J. Opt. Soc. Am. A* 6: 153-157.
9. Ohba, N. & Tanino, T. (1976) Unilateral colour vision defect resembling tritanopia. *Modern Problems in Ophthalmol.* 17: 331-335; Alpern, M., Kitahara, K. & Krantz, D.H. (1983). Perception of colour in unilateral tritanopia. *J. Physiol. (Lond.)* 335: 683-697.
10. Abramov, I., Gordon, J & Chan, H. (1991) Color appearance in the peripheral retina: Effects of stimulus size. *J. Opt. Soc. Am. A* 8: 404-414.
11. De Valois, R.L. & De Valois, K.K. (1993) A multi-stage color model. *Vision Res.* 33: 1053-1065.

## Spatial vision based upon color differences

Karen K. De Valois

University of California at Berkeley, Vision Science and Psychology  
Berkeley, California 94720

### ABSTRACT

The role of color vision is not limited to the acquisition and appreciation of information about the spectral composition of stimulus patches, its historical realm. Rather, color vision allows one to use information about stimulus spectral parameters to determine other interesting and relevant object characteristics. To understand the role of color in spatial vision, it is necessary to examine both the extent to which spatial discriminations can be based solely upon color differences and the interaction between color and luminance variations when they are simultaneously present.

The well-known differences in the spatial and temporal contrast sensitivity functions for color and luminance and the apparently impoverished input from the color mechanisms to certain higher functions obscure the fact that spatial discriminations based solely upon color differences are quite good. For example, spatial frequency discriminations between high-contrast patterns at isoluminance are only slightly poorer than for comparable luminance patterns, averaging about 5-6% of the base frequency. Similarly, orientation differences of about 1 deg between isoluminant patterns can be reliably discriminated at high contrasts, even for stimuli that lie along a tritanopic confusion axis<sup>45</sup>. Similar comparisons from several tasks will be reviewed, as will tasks (e.g., masking and adaptation) involving color-luminance interactions. These provide information about the target behavior that must ultimately be explained if the physiological basis of color vision is to be understood.

### 1. INTRODUCTION

That color lends beauty to the world cannot be denied. Indeed, color alone, like that we see lighting the evening sky, is a stimulus of such power and loveliness that it draws our eyes and inspires artists, even though it may define no object and carry no particular meaning. Yet the very potency of color and the enormous neural investment dedicated to its processing argue that it must fill other roles than the appreciation of sunsets, as profound as that may be. To understand what functions color vision may support, one must first know what vision is like when only color differences are present, and when both color differences and effective intensity differences are present in the scene. Only then can we begin to determine how we use information about stimulus spectral parameters to determine other interesting and relevant characteristics of the visual world.

Color contrast in the absence of luminance contrast can be produced (with great care) in the laboratory, but it rarely occurs in nature. It is nonetheless of some interest to determine the characteristics of vision in this condition of isoluminance, since it defines one limiting condition. Much of the classical color discrimination literature represents performance at isoluminance. For example, purity discrimination functions have traditionally been measured with luminance-equated stimuli<sup>32</sup>. These measurements, however, have virtually always been made either with spatially unpatterned stimuli (a 2° field) or with a simple bipartite field. Since the interest has most often been in the analysis and encoding of information about color itself, this is an appropriate way to approach the problem. If, however, one's interest is in the possible role of color differences in spatial vision, then other stimuli and other experimental methods are called for. Several of the studies briefly described below represent attempts to discern the extent to which color vision can support useful spatial vision.

## 2. SPATIAL VISION AT ISOLUMINANCE

### 2.1 Pattern detection

How we analyze and encode information about color *per se*--that is, about the spectral distribution of the light arriving from a particular point in space--is a very different question from that of how we use information about spectral differences to determine the characteristics of objects in the environment. The modern era of research into the role of color in spatial vision began with the pioneering studies of de Lange<sup>26,27</sup>, who introduced the techniques of linear systems analysis to vision. A few years later, van der Horst & Bouman<sup>18,19</sup> published spatial contrast sensitivity functions (CSF) for isoluminant chromatic gratings. Their results were the first to show the fundamental ways in which chromatic contrast sensitivity differs from luminance contrast sensitivity. They found that the spatial CSF for color falls off much sooner on the high spatial frequency end, and that unlike the luminance CSF, the chromatic CSF shows no low-frequency attenuation. At the lowest spatial frequencies they used (0.7 c/deg), chromatic contrast sensitivity was at its highest level. Over a range of spatial frequencies contrast sensitivity was flat, then dropped rapidly as frequency continued to increase. Although many others have measured the chromatic spatial CSF since<sup>13,35</sup>, often with better stimulus control and experimental methodology, the basic results of van der Horst and Bouman stand. Although the range of spatial frequencies that can be detected using color differences alone is apparently restricted in comparison to the range that can be analyzed using intensity differences, it is nonetheless of significant breadth. Geissler<sup>11</sup>, by use of an ideal detector model, has argued that much of the difference between the CSFs for color and for luminance can be accounted for solely by consideration of receptoral and pre-receptoral factors. This strongly suggests that the neural processing substrates for chromatic and luminance CSFs are equally efficient.

### 2.2 Pattern discrimination

Merely detecting the presence of a pattern is of little use, however. To be of practical value, a visual system should be able to discriminate readily between different objects. Accordingly, it is of interest to know that color differences alone are sufficient to support very good discriminations of either spatial frequency or orientation, even in comparison to performance with luminance-varying gratings<sup>2,4,36,43</sup>. Webster, De Valois & Switkes<sup>45</sup> found that spatial frequency differences of about 4-7% and orientation differences of a degree or so can be reliably discriminated in high-contrast isoluminant gratings. It is particularly interesting to note that these spatial discriminations can even be made with patterns that are defined solely by variations along a tritanopic confusion axis. The S-cone dependent system, although it is often considered to be inferior in several ways, is capable of supporting quite reasonable spatial discriminations. It is also especially significant that orientation discrimination can be performed well at isoluminance because it has been reported that striate cortical neurons that are sensitive to and selective for color differences in the absence of luminance differences have little or no orientation selectivity<sup>27</sup>. The lack of behavioral orientation sensitivity at isoluminance, if such were found, would be damning to the suggestion that color vision can support reasonable spatial vision.

One way to assess the tuning of the color system for spatial frequency and orientation is by the use of pattern-selective adaptation. As Gilinsky<sup>12</sup> and Blakemore and Campbell<sup>1</sup> first showed, prolonged adaptation to a high-contrast grating of a particular spatial frequency and orientation produces a transient but large loss in contrast sensitivity for test gratings that are spatially similar. It is presumed that the bandwidth of the adaptation effect is related (though probably not simply) to the bandwidths of the underlying adapted mechanisms. Bradley, Switkes and De Valois<sup>3</sup> found that the spatial-frequency bandwidth of the contrast sensitivity loss following adaptation to an isoluminant grating is quite similar to that seen for adaptation to a luminance-varying grating, while the orientation bandwidth for color is somewhat broader than the corresponding bandwidth for luminance. This is significant because the presence of a frequency band-limited adaptation effect, like the frequency-selective masking described below, implies that isoluminant stimuli are analyzed by frequency-selective channels similar to those responsible for the analysis of luminance patterns. To the extent that the potential fineness of spatial analysis is related to the selectivity of the underlying channels, these results suggest that the color vision system is only slightly inferior to luminance vision in its ability to analyze spatial patterns.

Another method of estimating the degree of selectivity for spatial frequency is by using a masking paradigm. The presence of a high-contrast grating of a given spatial frequency makes it more difficult to detect a low-contrast test grating of similar frequency and orientation. As with selective adaptation effects, the bandwidths measured using masking are quite similar for color-varying and luminance-varying test and adaptation gratings<sup>8,42,28,41</sup>. Both adaptation and masking studies, thus, suggest that color-varying patterns are analyzed by detectors with spatial selectivity that is very much like that in the mechanisms responsible for the analysis of luminance-varying patterns. Masking studies differ from adaptation studies in one interesting respect, however. The presence of a chromatically-varying mask can greatly impede the detection of a similar and coextensive luminance-varying test<sup>6,8,42</sup>. The converse is not found. A luminance-varying mask has little effect on the detectability of a superimposed chromatic test. When a selective adaptation paradigm is used, however, very little interaction between luminance and chromatic variations are found<sup>3</sup>.

Another question of interest is whether color differences alone are adequate to support the determination of depth in a visual pattern. There are many cues to depth, both monocular and binocular. Binocular disparity, the small differences in the retinal projection of a single object in the two eyes, underlies stereopsis, one of the finest of all visual spatial abilities. Although stereopsis is present and used in most visual patterns, it is difficult to study in isolation without the intrusion of monocular depth cues. One commonly-used way of assessing the binocular cue of stereopsis is by the use of random-dot stereograms<sup>21</sup>, in which the monocular information alone does not allow the observer to determine depth. Although there have been studies of random-dot stereopsis at isoluminance, there has not been agreement on the conclusions. Lu and Fender<sup>31</sup> found that color differences alone could not support stereopsis. De Weert and Sadza<sup>46</sup>, however, found that their subjects were able to successfully judge relative depth based on isoluminant random-dot stereograms. They used forced-choice psychophysical methods and trained their subjects for long periods. Even then, they report that although subjects can make accurate judgments, the perception of depth in isoluminant stereograms is strange and quite different from that seen in similar luminance-varying stereograms.

Reports are similarly mixed on the question of whether monocular perspective cues produce a sensation of depth at isoluminance. Livingstone and Hubel<sup>30</sup> report that they do not, but Troscianko, Montagnon, Le Clerc, Malbert and Chanteau<sup>44</sup> find that they do. In this, as with many similar questions, whether and how well a given task can be solved at isoluminance may depend in large part upon how the question is structured and what methods are used to assess performance.

Another fundamental task of spatial vision is to determine the relative positions of different objects. One positional task is that used to assess vernier acuity, in which a subject is asked to determine, for example, whether one stimulus is to the left or to the right of another that is displayed above it. When the object differs from its background in luminance, performance can be very good indeed. Thresholds of a few seconds of arc are commonly found and are largely independent of the spatial form of the stimuli that are compared<sup>47</sup>. Many of the stimuli used to study vernier acuity for luminance-varying patterns are not appropriate for use at isoluminance, both because of the problem of chromatic aberration and because of the reduced contrast sensitivity at high spatial frequencies. When appropriate stimuli are used, however, and when they are equated in terms of detectability, it is found that vernier acuity at isoluminance is very similar to that found with similar luminance stimuli<sup>24,25</sup>. Surprisingly, there is no significant loss in positional sensitivity for patterns that vary only in color.

In summary, when the task is structured appropriately, color differences are capable of supporting quite good spatial vision. In terms of both detection and discrimination, performance on a variety of tests is surprisingly good when only color variations are present in the pattern. The data briefly described above are summarized in a table and presented at the end of this chapter.

### 3. COLOR-SELECTIVE INTENSITY ENCODING

It is important to determine how well color differences alone can be used to make spatial discriminations, and some of the relevant literature has been briefly reviewed above. None of these studies, however, addresses one question that may be of fundamental importance. It is clear from the physiological studies of retina, lateral geniculate and striate cortex that nearly all neurons that respond well to color differences also respond quite well to intensity differences in a non-color-varying stimulus. That these neurons cannot properly be described as encoding luminance contrast *per se* is obvious from their non- $V_\lambda$  spectral sensitivity functions, yet we know remarkably little about the effect of stimulus chromaticity on their responses to intensity variations.

To what extent the intensity-coding responses of such cells provide a significant input to vision is not known. It is widely assumed, for example, that most such neurons do not participate in the detection of luminance-varying stimuli at threshold contrasts<sup>22,39</sup>, although there is considerable disagreement and some compelling evidence to the contrary. Merigan<sup>34</sup> and Schiller, Logothetis & Charles<sup>38</sup> have measured the effects on luminance contrast sensitivity of the selective ablation of either parvocellular or magnocellular pathway components. Both groups conclude that the largely color-opponent parvocellular pathway neurons are at least partially responsible for the detection at threshold of luminance-varying gratings.

Further evidence for the involvement of color-selective neurons in the threshold detection of intensity variations comes from studies of hue recognition. Finkelstein & Hood<sup>10,17</sup> have studied the detection of an intensity increment upon a background. They found that at threshold luminance contrasts, the hue of the stimulus spot can be accurately reported. Using a different paradigm, Guth and his coworkers<sup>14,15</sup> tested the applicability of Abney's Law to the detection of a luminance increment upon a dark background. Abney's Law states that luminances are additive, irrespective of the spectral distribution of the component lights. Guth found, however, that for certain wavelength combinations, additivity failed dramatically. A half-threshold amount of a red, for example, cannot be added to a half-threshold amount of green to produce a detectable amount of yellow. The subadditive behavior observed can be most readily explained by assuming that detection is subserved by spectrally opponent mechanisms like those found in retina and LGN.

Another class of experiments that suggest the involvement of color-selective intensity coding mechanisms is the color-contingent aftereffects. McCullough<sup>33</sup> had subjects adapt to two alternating luminance-varying gratings of different colors. If the adaptation patterns were, say, a red-black vertical grating and a green-black horizontal grating (presented alternately in the same retinal location), the perceived hue of a subsequently-viewed white-black grating depended upon its orientation. If the test grating were vertical, it would appear to be (a desaturated) green-black; if it were horizontal, it would appear red-black. The hue of the test grating appeared to be complementary to the hue of the adaptation grating of the same orientation. This demonstration was particularly significant because both adaptation (as well as test) gratings were presented in the same retinal location. Thus, differential receptor adaptation could not account for the effect on perceived hue. A similar effect was found by Hepler<sup>16</sup>, who produced directionally-selective color-contingent motion aftereffects.

Perhaps the most interesting observation comes from a particularly well-studied clinical case presented by Hyvarinen and Rovamo<sup>20,37</sup>. Their patient, a diabetic woman with epilepsy, experienced a medical crisis during which she became nearly blind to luminance contrast. Over a period of a few months her visual behavior became more nearly normal, but one striking exception remained. She was virtually completely unable to detect luminance contrast in achromatic stimuli. When measured psychophysically using forced-choice methods, and when assessed using visually evoked potentials, she gave no evidence of having useful vision for black-white gratings. However, the same test stimuli were detected with almost normal sensitivity if the subject viewed them through colored lenses. Note that in this case the detection of the patterns was not based upon the detection of color variations in the stimuli. There were none, whether the test gratings were white-black or bright red-dark red. The task in each case was the detection of a luminance variation. What determined detectability, however, was not only the level of luminance contrast, but also the mean chromaticity of the stimulus.

Each of these observations, although they are quite diverse in character, suggests the important involvement of color-selective mechanisms in the encoding of information about intensity variations in stimuli. They do not necessarily imply joint sensitivity to both color and intensity variations in single multiplexing neurons, although that would certainly be an easy and obvious way to accomplish these tasks. Color-contingent pattern adaptation, for example, might result from the network properties of a complex of cells, each of which is more restricted in its stimulus selectivity. It does seem unlikely, however, that the visual system would first develop exactly the sorts of neurons that could easily multiplex information about both color and intensity variations, then discard them only to recreate the same ability later through complex networks.

Why would it be of interest and importance to have color-coded information about intensity variations? Two reasons come to mind. The first has been discussed before<sup>7</sup>. The analysis of color variations is a particularly good way to discriminate between luminance contrast produced by shadows and contrast produced by changes in reflectance, that is, object borders. The images of most objects differ from their surroundings in both spectral distribution and effective intensity (or luminance). An edge that is defined by both color difference and luminance difference is likely to be associated with an object border. An edge that is formed by luminance change alone (or almost so) is more likely to indicate a shadow border, information that may be quite useful in determining object shape<sup>5</sup>. Edges produced by shadows can have extremely high luminance contrast, but it is much rarer that they are high in chromatic contrast. If the visual system must first segregate the various parts of a complex stimulus into different objects, as opposed to determine shape *per se*, the use of correlated chromatic/intensity information is a useful way to begin. That we do use correlated chromatic/intensity information to segregate objects in a complex field is shown by studies of sliding and coherence in moving plaid patterns<sup>23</sup>. Whether an orthogonal luminance plaid pattern is seen as coherent or transparent can be determined in large part by the way in which chromatic variation is added to its components. If the same chromatic pattern is added to the two components in the same relative phase, coherence will not be disrupted. However, if identical chromatic variation is added again, *but in opposite phases with respect to the two luminance gratings*, the plaid will appear to break apart. The two oriented gratings will appear to move independently in visual transparency.

A second way in which the joint analysis of chromatic and luminance variations could be particularly useful is in the discrimination between changes in the illuminant and changes in the object. If a bright red circle upon a dark red background is replaced by an otherwise identical bright green circle upon a dark green background, it will not be possible for an observer to determine whether the object and background have changed or the illuminant has changed. If the red circle is seen upon a grey background, however, and it is replaced by a green circle upon the same grey background, it will be obvious that the object, not the general illuminant, has changed. Although dramatic changes of illuminant from red to green do not occur in nature, the color of the sky does change significantly, if gradually, over the course of a day. The absolute spectral distribution of the light reflected from a red flower is not constant across the hours, but it will remain redder than its surrounding foliage regardless. Should it appear red at noon and blue in late afternoon, one would suspect a substitution, not just a change in the illumination. Thus, although it is not necessary to identify all the flower's spatial characteristics (its fine structure, for example) based on color differences, correlating the luminance-defined structure with a crudely-defined color variation can aid in maintaining object constancy and correctly identifying it as the same object under a variety of viewing conditions.

## S. ACKNOWLEDGEMENTS

The preparation of this chapter and much of the work reported herein were supported by grants from the National Eye Institute (EY00014) and the National Science Foundation (BNS 88 19867).

	LUMINANCE	COLOR
Spatial CSF	band-pass higher high-frequency cut (van der Horst & Bouman, 1969; Mullen, 1985)	low-pass lower high-frequency cut (van der Horst & Bouman, 1969; Mullen, 1985)
Spatial Freq. Discrimination	.02-.04f (Regan, et al., 1982; Thomas, 1983; Caelli, et al., 1983; Skottun, et al., 1987)	.04-.05f (Webster, et al., 1990)
Orientation Discrimination	0.65 deg or less (Eisner & MacLeod, 1980; Regan, et al., 1982; Bradley, et al., 1987; Webster, et al., 1990)	~ 1 deg (Webster, et al., 1990)
Contrast Discrimination	dipper function band-limited masking (Legge & Foley, 1980; Swift & Smith, 1983)	dipper function band-limited masking (Switkes, et al., 1988)
Grating Adaptation	band-limited in spatial freq. & orientation (Blakemore & Campbell, 1969)	band-limited in spatial freq. & orientation (Bradley, et al., 1988)
Random Dot Stereopsis	present (Julesz, 1971)	mixed data no - Lu & Fender, 1972 yes - De Weert & Sadza, 1983
Perspective Depth	present	mixed data no - Livingstone & Hubel, 1987 yes - Troscianko, et al., 1991
Vernier Acuity	excellent for a variety of stimuli (Westheimer & McKee, 1977; Krauskopf & Farell, 1991)	comparable for equally detectable Gabor patterns (Krauskopf & Farell, 1991; Kooi, et al., 1991)

#### 4. REFERENCES

1. C. Blakemore and F. Campbell, "On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images," *J. Physiol. (Lond.)* 203, pp. 237-260, 1969.
2. A. Bradley, B. Skottun, I. Ohzawa, G. Sclar and R. Freeman, "Visual orientation and spatial frequency discrimination: A comparison of single neurons and behavior," *J. Neurophysiol.* 57, 755-772, 1987.
3. A. Bradley, E. Switkes and K. K. De Valois, "Orientation and spatial frequency selectivity of adaptation to colour and luminance gratings," *Vision Res.* 28, 841-856, 1988.
4. T. Caelli, H. Brettel, I. Rentschler and R. Hilz, "Discrimination thresholds in the two-dimensional spatial frequency domain," *Vision Res.* 23, 129-133, 1983.
5. P. Cavanagh and Y. Leclerc, "Shape from shadows," *J. exp. Psychol.: Human. Percep. Perform.* 15, 3-27, 1989.
6. G. R. Cole, C. F. Stromeyer, III, R. E. Kronauer, "Visual interactions with luminance and chromatic stimuli," *J. O.S.A. A* 7, 128-140, 1990.
7. K. K. De Valois and F. L. Kooi, "The role of color in spatial vision," in L. Harris and M. Jenkin, (Eds.), *Spatial Vision in Humans and Robots*, New York: Cambridge University Press, in press.
8. K. K. De Valois and E. Switkes, "Simultaneous masking interactions between chromatic and luminance gatings," *J.O.S.A.* 73, 11-18, 1983.
9. A. Eisner and D. I. A. Macleod, "Blue-sensitive cones do not contribute to luminance," *J. O. S. A.* 70, 121-123, 1980.
10. M. A. Finkelstein and D. C. Hood, "Opponent-color cells can influence detection of small, brief lights," *Vision Res.* 22, 89-95, 1982.
11. W. S. Geisler, "Sequential ideal observer analysis of visual discriminations," *Psychol. Rev.* 96, 267-314, 1989.
12. A. S. Gilinsky, "Orientation-specific effects of patterns of adapting light on visual acuity," *J.O.S.A.* 58, 13-18, 1968.
13. E. M. Granger and J. C. Heurtley, "Visual chromaticity-modulation transfer function," *J.O.S.A.* 63, 1173-1174, 1973.
14. S. L. Guth, "Nonadditivity and inhibition among chromatic luminances at threshold," *Vision Res.* 7, 319-327, 1967.
15. S. L. Guth, N. J. Donley and R. T. Marrocco, "On luminance additivity and related topics," *Vision Res.* 9, 537-575, 1969.
16. N. Hepler, "Color: a motion-contingent aftereffect," *Science* 152, 376-377, 1968.

17. D. C. Hood and M. A. Finkelstein. "A case for the revision of textbook models of color vision: the detection and appearance of small brief lights," in *Colour Vision: Physiology and Psychophysics*, J. D. Mollon and L. T. Sharpe (Eds.), pp. 385-398, Academic Press, London, 1983.
18. G. J. C. van der Horst and M. A. Bouman, "Spatiotemporal chromaticity discrimination," *J.O.S.A.* 59, 1482-1488, 1969.
19. G. J. C. van der Horst, C. M. M. de Weert and M. A. Bouman, "Transfer of spatial chromaticity-contrast at threshold in the human eye," *J.O.S.A.* 57, 1260-1266, 1967.
20. L. Hyvarinen and J. Rovamo, "Acquired blindness for achromatic stimuli," *Doc. Ophthalmol. Proc. Ser.* Vol 30. 94-99, 1981.
21. B. Julesz, *Foundation of Cyclopean Perception*, University of Chicago Press, Chicago, 1971.
22. E. Kaplan, B. B. Lee and R. M. Shapley, "New views of primate retinal function," *Prog. Retinal Res.* 9, 273-336, 1990.
23. F. L. Kooi, K. K. De Valois, E. Switkes and D. G. Grosof, "High-order factors influencing the perception of sliding and coherence of a plaid," *Perception* 21, 583-598, 1992.
24. F. L. Kooi, R. L. De Valois and E. Switkes, "Spatial localization across channels," *Vision Res.* 31, 1627-1631, 1991.
25. J. Krauskopf and B. Farell, "Vernier acuity: effects of chromatic content, blur and contrast," *Vision Res.* 31, 735-749, 1991.
26. H. de Lange, "Research into the dynamic nature of the human fovea→cortex systems with intermittent and modulated light: I. Attenuation characteristics with white and colored light," *J. O. S. A.* 48, 777-784, 1958.
27. H. de Lange, "Research into the dynamic nature of the human fovea→cortex systems with intermittent and modulated light: II. Phase shift in brightness and delay in color perception," *J. O. S. A.* 48, 784-789, 1958.
28. G. E. Legge and J. M. Foley, "Contrast masking of human vision," *J.O.S.A.* 70, 1458-1471, 1980.
29. M. S. Livingstone and D. H. Hubel, "Anatomy and physiology of a color system in the primate visual cortex," *J. Neurosci.* 4, 309-356, 1984.
30. M. S. Livingstone and D. H. Hubel, "Psychophysical evidence for separate channels for the perception of form, color, movement, and depth," *J. Neurosci.* 7, 3416-3468, 1987.
31. C. Lu and D. H. Fender, "The interaction of color and luminance in stereoscopic vision," *Invest. Ophthalmol. Vis. Sci.* 11, 482-490, 1972.
32. L. C. Martin, F. L. Warburton and W. J. Morgan, "Determination of the sensitiveness of the eye to differences in the saturation of colours," *Med. Res. Council, Rept.* 188, London, 1933.
33. C. McCollough, "Color adaptation of edge-detectors in the human visual system," *Science* 149, 1115-1116, 1965.

34. W. H. Merigan, "Chromatic and achromatic vision of macaques: role of the P pathway," *J. Neurosci.* 9, 776-783, 1989.

35. K. T. Mullen, "The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings," *J. Physiol. (Lond.)* 359, 382-400, 1985.

36. D. Regan, S. Bartol, T. J. Murray and K. I. Beverly, "Spatial frequency discrimination in normal vision and in patients with multiple sclerosis," *Brain* 105, 734-754, 1982.

37. J. Rovamo, L. Hyvarinen and R. Hari, "Human vision without luminance-contrast system: selective recovery of the red-green colour-contrast system from acquired blindness," *Docum. Ophthal. Proc. Series, Vol. 33,* pp. 457-466, 1982.

38. P. H. Schiller, N. K. Logothetis and E. R. Charles, "Role of the color-opponent and broad-band channels in vision," *Visual Neurosci.* 5, 321-346, 1990.

39. R. M. Shapley, "Visual sensitivity and parallel retinocortical channels," *Ann. Rev. Psychol.* 41, 635-658, 1990.

40. B. C. Skottun, A. Bradle, G. Sclar, I. Ohzawa and R. D. Freeman, "Thee effects of contrast on visual orientation and spatial frequency discrimination: A comaprison of single cells and behavior," *J. Neurophysiol.* 57, 773-786.

41. D. J. Swift and R. A. Smith, "Spatial frequency masking and Weber's law," *Vision Res.* 23, 495-505, 1983.

42. E. Switkes, A. Bradley and K. K. De Valois, "Contrast dependence and mechanisms of masking interactions among chromatic and luminance gratings," *J.O.S.A. A* 5, 1149-1162, 1988.

43. J. P. Thomas, "Underlying psychometric function for detecting gratings and identifying spatial frequency," *J.O.S.A. A* 7, 751-757, 1983.

44. T. Troscianko, R. Montagnon, J. Le Clerc, E. Malbert and P. L. Chantcau, "The role of colour as a monocular depth cue," *Vision Res.* 31, 1923-1929, 1991.

45. M. A. Webster, K. K. De Valois and E. Switkes, "Orientation and spatial frequency discrimination for luminance and chromatic gratings," *J.O.S.A. A* 7, 1034-1049, 1990.

46. C. M. M. de Weert and K. J. Sadza, "New data concerning the contribution of colour differences to stereopsis," in *Colour Vision: Physiology and Psychophysics*, J. D. Mollon and L. T. Sharpe (Eds.), pp. 553-562, Academic Press, New York , 1983.

47. G. Westheimer and S. P. McKee, "Spatial configurations for hyperacuity," *Vision Res.* 17, 941-947, 1977.

# Neurobiological mechanisms of cortical direction selectivity

Curtis L. Baker, Jr.<sup>#</sup>, and Jane C. Boulton\*

\*McGill Vision Research Unit, Dept. of Ophthalmology, 687 Pine Ave. W., H4-14, Montreal, Quebec, Canada H3A 1A1

<sup>#</sup>Utrecht Biophysics Research Institute, Utrecht, The Netherlands

## ABSTRACT

From consideration of a number of types of apparently linear and nonlinear behavior of direction selectivity of visual cortex neurons, it will be argued that there are at least two fundamentally different types of motion computation. The first, designated "quasi-linear", entails a summation of afferent signals which are in approximate quadrature phase, both spatially and temporally (e.g., "lagged" and "non-lagged" LGN afferents, in the cat); the summation may be of a linear or a partially nonlinear nature, but is carried out on specific signals falling within a relatively restricted spatial frequency passband and confined receptive field. The second, referred to as "nonlinear", involves a highly nonlinear integration of additional, non-specific afferent signals, generally outside the conventional spatiotemporal frequency passband of a neuron, and also outside of the "classical" receptive field.

Some novel aspects of this formulation are: the same neuron may exhibit both quasi-linear and nonlinear behavior; quasi-linear mechanisms may display substantial nonlinearities, possibly accounting for detection of some "non-Fourier" stimuli. Data are presented to illustrate the idea that white noise analysis methods are well-suited to characterize the spatiotemporal nonlinearities of "quasi-linear" mechanisms, but fail to provide insight into the processing of "nonlinear" mechanisms.

## INTRODUCTION

The first two sections of this paper will provide an overview of some basic concepts in visual neurophysiology and in modeling of visual receptive fields, which will be relevant background for the subsequent discussion of mechanisms of cortical direction selectivity.

The transition from the lateral geniculate nucleus (LGN) to the visual cortex represents a major transformation in visual processing - it is here that a high divergence gives rise to a rich representation in terms of neurons selective for local orientation, spatial frequency, disparity, and velocity components of the image. This rich representation provides the fundamental "basis set" upon which more complex computations such as optic flow pattern analysis can be built.

Visual cortex cells are very different from those of the lateral geniculate nucleus: they are remarkably selective for orientation and spatial frequency, not responding at all to a spatially uniform field. Also in contrast to the LGN, these neurons fire only in response to some kind of temporal modulation, and usually are selective to temporal frequency for sinusoidally contrast-modulated stimuli. They also usually show a selectivity for direction and speed of motion, which will be the principal subject of this paper. The visual cortex of the cat is a good choice for the study of cortical motion detection - nearly all of the neurons encountered in the early thalamo-recipient areas (areas 17 and 18) are at least partially, and often entirely, direction-selective.

The following set of assumptions are largely implicit in this work: (1) direction selectivity (as well as spatial frequency and orientation tuning) are first formed crudely from LGN afferents, probably in layer IV (see, e.g., Saul and Humphrey<sup>1</sup>); (2) this crude selectivity is subsequently sharpened up by inhibitory interactions between differently-tuned

cells (Eysel et al<sup>2</sup>); (3) many cells may entirely "inherit" their tuning properties from LGN-recipient cortical neurons. The agenda of this paper is aimed at the primary process (1) - the genesis of direction selectivity.

This discussion will be primarily concerned with cortical neurons having "simple" type receptive field organization - segregated, elongated regions of alternating excitatory (ON) and inhibitory (OFF) response. A simple cell's orientation-preference for visual stimuli corresponds to its direction of elongation, and its optimal spatial frequency for sinewave gratings corresponds to the periodicity of the ON and OFF-responding zones. However the nature of direction selectivity in cortical neurons has generally not been found to correspond to classical receptive field spatial properties. The preferred direction of motion is not systematically related to, e.g., asymmetries in the spatial layout of receptive field regions, and direction selectivity can be demonstrated with moving bars whose total excursion is entirely confined within ON or OFF zones (Goodwin, Henry, and Bishop<sup>3</sup>).

"Delayed comparison" models have been proposed for explaining direction selectivity in other biological systems - mammalian retinal ganglion cells (Barlow and Levick<sup>4</sup>) and the fly visual system (Reichardt<sup>5</sup>). One example of such a model is illustrated in Fig.6A - signals from adjacent visual field locations, separated by a distance  $D_0$ , are compared by an "AND-gate" operator; the signal from one position is temporally delayed by a time  $T_0$ . The logical condition for the AND-gate to respond is met only by a visual stimulus moving from left to right, at a speed of  $D_0 / T_0$ . Some fundamental problems with this type of model for visual cortex neurons, such as the lack of front-end spatial filtering (e.g., allowing spatial "aliasing") and undesirable temporal frequency dependence of the fixed delay, can be fixed by the addition of appropriate spatial and temporal filters (e.g., van Santen and Sperling<sup>6</sup>). This early emphasis on models using nonlinear, AND or AND-NOT gating has been overtaken in popularity by a seemingly different kind of linear model.

## MODELS OF DIRECTION SELECTIVITY BASED ON LINEAR FILTERING

The currently most popular kind of models for cortical neuronal direction selectivity are those based on linear spatiotemporal filtering. Only a brief overview of the basic ideas will be presented here - see, e.g., Adelson and Bergen<sup>7</sup> and Heeger<sup>8</sup> for a more extensive introduction.

A visual stimulus moving in the frontoparallel plane is in general an intensity function of two spatial dimensions and of time. A useful simplification is to ignore the spatial dimension orthogonal to the direction of motion, giving a function which is conveniently represented as an intensity plot. For example, Fig.1A-C shows such "space-time diagrams" for a light bar moving rightwards at three successively increasing speeds. Some other examples: a stationary bar would be a vertical stripe; a spatially uniform field whose intensity is sinusoidally modulated would be a horizontal grating; and a drifting grating would be a tilted grating whose orientation corresponds to the velocity of movement.

Spatiotemporal filters may also be represented with intensity plots as functions of spatial position (abscissa) and time (ordinate); for example, Fig. 1D shows a "space-time oriented" filter which responds maximally to a stimulus moving rightwards at a speed corresponding to that of the stimulus in the middle panel of Fig. 1B. Intuitively, this spatiotemporal filter can be thought of as a neuronal receptive field, but in space-time: those stimuli whose space-time history best "line up" (correlate) with the filter function will be those that maximally activate the neuron. Note in the example of Fig.1D, the filter would respond well to a rightwards moving black or white bar, or a rightwards moving sinewave grating, providing the velocity matched the filter's space-time orientation (and in the case of a grating, if the spatiotemporal frequency matched the spacing of the positive and negative regions in the filter function); but any of these stimuli drifting leftwards would be ineffective.

Formally, the response  $r(t)$  of a linear spatiotemporal filter is expressed as an integral :

$$r(t) = \iint k(\sigma, \tau) x(\sigma, \tau-t) d\sigma d\tau \quad (1)$$

where  $x(s,t) =$  the stimulus and  $k(s,t) =$  the filter (as functions of space,  $s$ , and time,  $t$ ). A defining characteristic of such a linear system is that it obeys the superposition principle: the response to the sum of two stimuli will be equal to the sum of responses to each of the stimuli alone. If the filter function were "space-time separable", it would obey the relationship:

$$k(\sigma, \tau) = k_s(\sigma) \cdot k_t(\tau) \quad (2)$$

where  $k_s(\sigma)$  is a spatial profile function (independent of time lag,  $\tau$ ) and  $k_t(\tau)$  is a temporal response function (independent of spatial position,  $\sigma$ ). Note that a space-time oriented filter, by definition, does not satisfy this relationship (ie, it is non-separable).

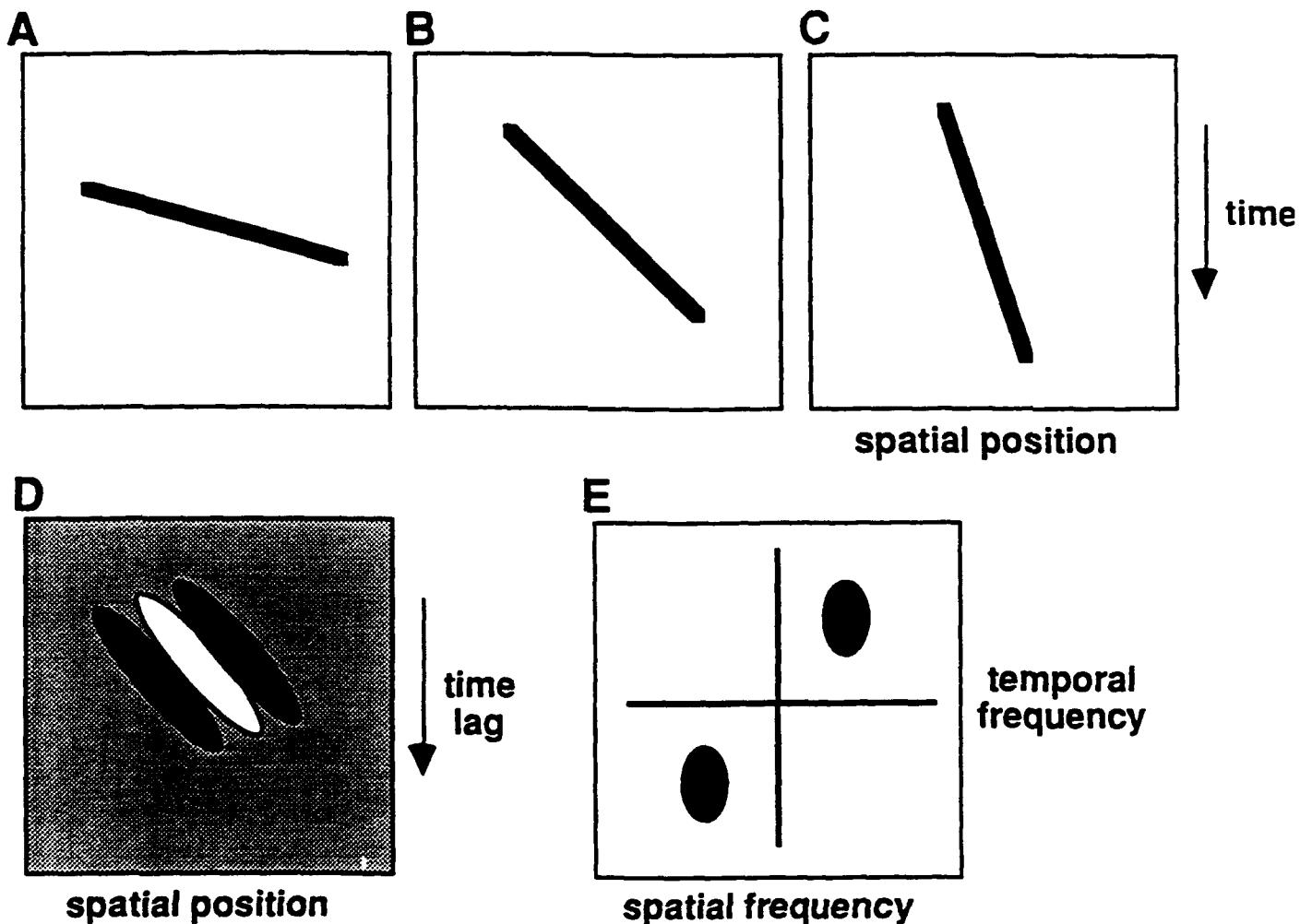


Fig.1 Space-time diagrams and linear spatiotemporally oriented filtering. A., B. C. Space-time diagrams of a bar stimulus moving rightwards at three successively increasing speeds. D. Spatiotemporal filter function, or "kernel", having oriented positive (white) and negative (black) regions. This filter would respond more strongly to the stimulus in B than those of A or C. D. Magnitude of Fourier transform of filter in C; note its bandpass nature, both in spatial and in temporal frequency.

To turn such a linear filter into a viable model of a cortical neuron, at least two assumptions are needed. Firstly, the filter response is taken to represent the average instantaneous frequency of the all-or-none impulses ("spikes") of neuronal response. Secondly, since cortical neurons generally have negligible spontaneous impulse frequency, the linear

filter must be followed by a half-wave rectification to preclude negative responses - a negative value of impulse frequency is undefined. Technically such a rectification is a nonlinearity, but may be considered a "trivial" nonlinearity - the basic selectivity properties of the filter are still understandable entirely in terms of the linear "front-end" spatiotemporal summation.

Variations of this kind of linear model were first proposed primarily to explain human motion perception, but may be considered as candidate models of single neuron directionality. Watson and Ahumada<sup>9</sup> described a purely linear model whose filter design emphasized "quadrature" properties (see below). The "Elaborated Reichardt Detector" of van Santen and Sperling<sup>6</sup> was similar to Fig.6A, but with bandpass spatial filters at the front end to prevent aliasing, temporal filters replacing the pure delay element, and a pure multiplier instead of an "AND-gate". The model of Adelson and Bergen<sup>7</sup> involved a summation of squared responses of quadrature-phase pairs of space-time oriented filters such as in Fig.1D. More recent models which specifically address cortical neuronal data also use space-time oriented filters, but followed by a "half-square" operator (half-wave rectification followed by squaring), and often incorporating a contrast gain control mechanism (Albrecht and Giesler<sup>10</sup>; Heeger<sup>8</sup>). Such a point-wise nonlinearity following the linear filter improves the quantitative direction selectivity (Albrecht & Giesler<sup>10</sup>; DeAngelis et al<sup>11</sup>), and also allows a net directional response to "two-flash apparent motion" stimuli (Baker, in preparation).

Though these models incorporate some nonlinearities, they all use front-end filters that are spatiotemporally linear. An important consequence of this linearity is that only stimuli whose spatio-temporal Fourier transform has power spectral components falling within the passband of the filter function will elicit responses. (This follows from taking the Fourier transform of both sides of Eq.1).

Such models provide an appealing explanation for velocity tuning and contrast-invariant direction selectivity, while also showing a bandpass response for spatial and temporal frequency of sinewave gratings. For example, Fig.1E shows the magnitude of the Fourier transform of the filter function in Fig.1D - note that it is bandpass in both spatial and temporal frequency, another typical property of visual cortex neurons. Also the function in Fig.1E shows a separable dependence on spatial and temporal frequency, i.e., the spatial frequency tuning curve does not change in shape with temporal frequency. This kind of separability is typical of visual cortex neurons (e.g., Tolhurst and Movshon<sup>12</sup>; Friend and Baker<sup>13</sup>), and follows naturally from spatiotemporal linear models provided the *envelope* of the kernel function,  $k(s,t)$ , is separable in space,  $s$ , and time,  $t$  (McLean and Palmer<sup>14</sup>).

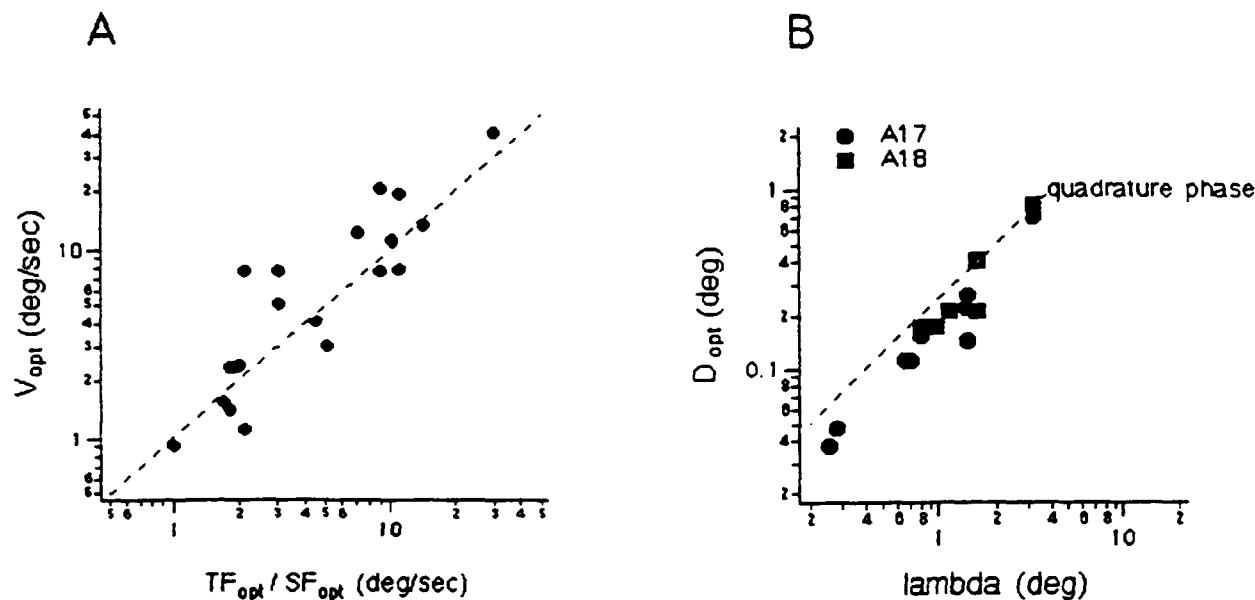
Further support for such models comes from experimental data based on sinusoidal grating contrast reversal data (Reid et al<sup>15</sup>; Albrecht and Giesler<sup>10</sup>; Tolhurst and Dean<sup>16</sup>) as well as "reverse correlation" studies which demonstrate that an estimate of the kernel function,  $k(s,t)$ , is spatiotemporally oriented in a manner which correlates with a given neuron's velocity tuning to bar stimuli (McLean and Palmer<sup>17</sup>) or with its bandpass tuning to sinewave gratings (DeAngelis et al<sup>11</sup>). Another quantitative prediction of such linear models is a lawful relationship between the optimal velocity,  $V_{opt}$ , for a bar stimulus, and the optimal spatial and temporal frequency,  $SF_{opt}$  and  $TF_{opt}$ , for drifting sinewave gratings; the scatterplot of Fig.2A, adapted from Baker<sup>18</sup>, shows that most cortical neurons adhere to the relationship:

$$V_{opt} = TF_{opt} / SF_{opt} \quad (3)$$

Different neurons have rather widely varying optimal spatial frequencies, and also somewhat of a variety of optimal temporal frequencies, providing a large "dynamic range" of possible optimal velocities across the population of cells. This range is further extended by a co-variation, such that neurons tuned to lower spatial frequencies tend to be tuned to higher temporal frequencies (Baker<sup>18</sup>; DeAngelis et al<sup>11</sup>).

An arguably optimal way of constructing spatiotemporally oriented filters is to build them from linear combinations of space-time separable filters, which are in quadrature spatial and temporal phase to one another (Adelson and Bergen<sup>7</sup>; Watson and Ahumada<sup>9</sup>). This idea is illustrated in Fig.3, which shows a kernel function approximating the one of Fig.1D, but constructed from separable inputs: kernel regions labeled N+ and N- represent approximations of filtering functions from short-latency ON- and OFF-centre LGN afferents, while L+ and L- denote longer-latency LGN afferents. In addition the N-filters have a biphasic (transient) temporal response, producing temporal reversals (NR) in the filter function.

Note that the N and L portions of the filter are in quadrature spatial phase (i.e., the peaks/valleys of one are aligned with the zero-crossings of the other) as well as approximate quadrature temporal phase.



**Figure 2.** Relationships between spatial and temporal parameters of receptive fields of visual cortex neurons; each symbol represents measurements from one neuron. A. The optimal velocity ( $V_{\text{opt}}$ ) for a moving bar stimulus is predicted (dashed line) from the ratio of optimal temporal to optimal spatial frequency, as measured with sinewave gratings (from Baker<sup>18</sup>). B. The optimal spatial displacement for multi-flash apparent motion falls slightly below quadrature spatial phase (dashed line) relation (from Baker, Friend, and Boulton<sup>20</sup>).

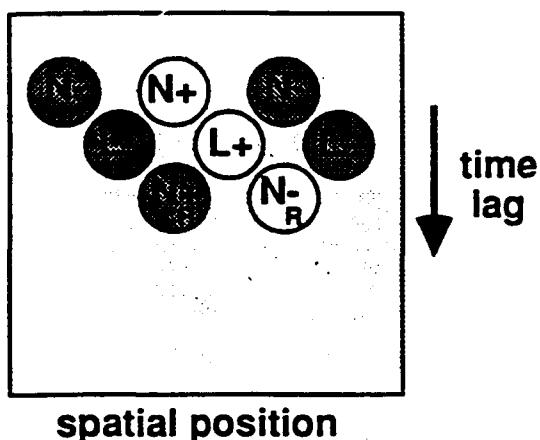
Support for such an approximate quadrature spatial phase relationship in cortical neurons comes from experiments using two-flash, "apparent motion" stimuli (Baker and Cynader<sup>19</sup>) as well as discretely jumping sinewave gratings (Fig.2B, from Baker, Friend, and Boulton<sup>20</sup>) - i.e., the optimal spatial displacement ( $D_{\text{opt}}$ ) to elicit direction selectivity is about (or often, slightly less than) one quarter of a spatial cycle ( $\lambda$ ) of the optimal spatial frequency for a given neuron:

$$D_{\text{opt}} \leq \lambda / 4 \quad (4)$$

This result is quantitatively consistent with human psychophysical studies of apparent motion with spatially band-limited stimuli, in which optimal motion aftereffects were obtained from jumping gratings (Baker, Baydala, and Zeitouni<sup>21</sup>), and optimal performance was obtained from 2-flash Gabor function kinematograms (Boulton and Baker<sup>22</sup>), for spatial displacements at or slightly less than a quarter of a cycle of the stimulus spatial frequency.

An important question is how such filters might be realized biologically. One appealing candidate substrate are "lagged" type LGN neurons (Mastronarde<sup>23</sup>), which have a greater latency to visual stimuli and a very roughly quadrature temporal phase relationship to "non-lagged" LGN cells (Saul and Humphrey<sup>24</sup>). If such LGN afferents were combined in spatial quadrature phase, a reasonable approximation to space-time oriented filtering could be achieved (as in Fig.3). Another worry lies in how to achieve linear behavior from inherently half-wave rectified LGN afferents - a commonly proposed solution is to combine oppositely signed (i.e., ON and OFF) signals in a "push-pull" arrangement, so that an excitatory part of the kernel function would be realized by an excitatory connection from ON-centre afferents combined with

inhibitory input from OFF-centre afferents. However, given the inherent nonlinearities of membrane biophysical mechanisms, it would be naive to expect entirely linear behavior.



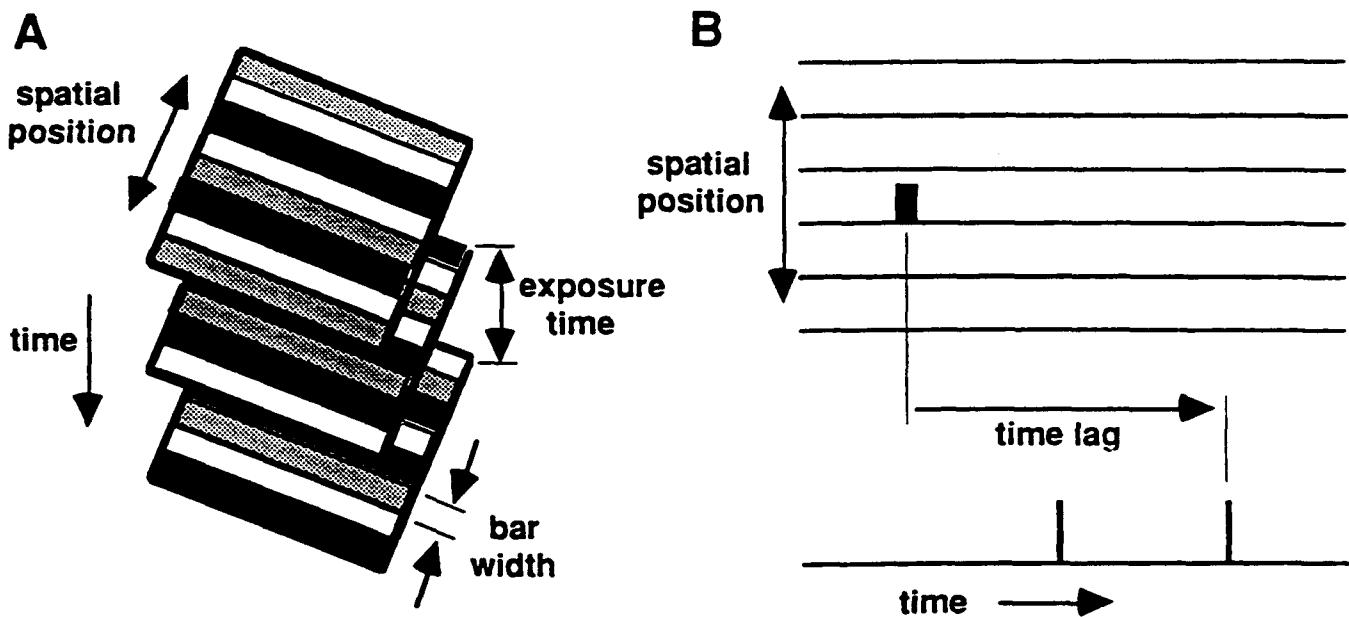
**Fig.3** Possible biological implementation of spatiotemporally oriented filter, from a sum of discrete space-time separable filters. N denotes "non-lagged" LGN afferents, L denotes "lagged" LGN afferents; + or - signs, and light or dark color, signify ON or OFF receptive field regions. R-subscript refers to sign-reversal due to biphasic temporal response. Note that Lagged filters are in approximate spatiotemporal quadrature phase relative to Non-lagged filters.

## White Noise Analysis

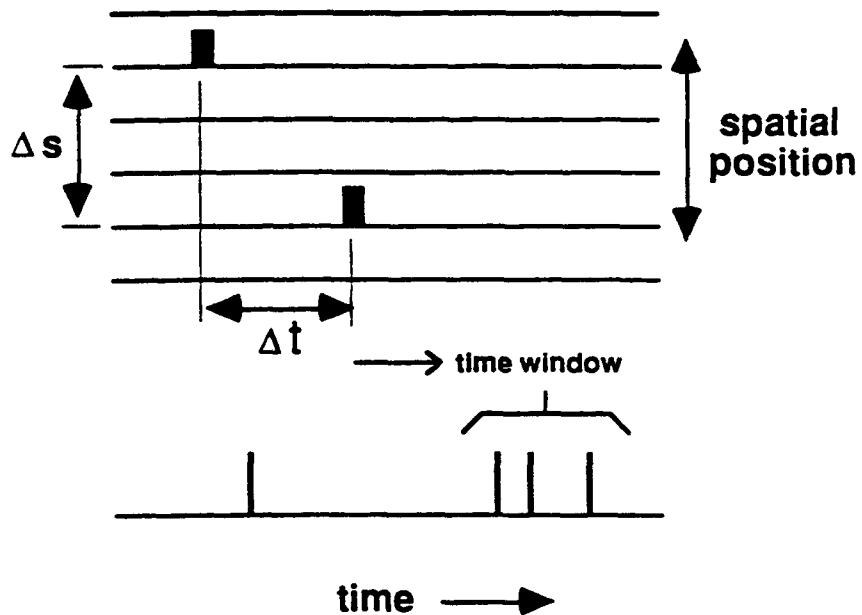
One potentially powerful approach for probing spatiotemporally linear and nonlinear properties relevant to direction selectivity is white noise analysis (Marmarelis and Marmarelis<sup>25</sup>). The essence of such a "system identification" technique is to employ a stimulus which is rich in frequencies, or equivalently, rich in spatial and temporal separations; such a mixture provides the opportunity to observe nonlinear interactions between, e.g., pairs of frequencies or of space-time positions. The stimulus used here is ternary white noise (Fig.4A): bar-shaped stimuli at each of 32 contiguous spatial positions are randomly made white, gray, or black with equal probability, and this random selection is re-chosen for each "exposure" time. Such an intensity distribution naturally maps onto the "ON-OFF" nature of neuronal receptive fields, as well as enabling use of a faster data analysis algorithm. In order to simplify estimation of nonlinear interactions, this stimulus is not an "m-sequence" (Sutter<sup>26</sup>). This stimulus corresponds closely to that used by Emerson et al<sup>27</sup>, but differs from those used by McLean and Palmer<sup>17</sup> and DeAngelis et al<sup>11</sup>, which presented only *single* bars or spots on each frame.

The neuronal response is cross-correlated with the white noise sequence to produce an estimate of the first kernel function, corresponding to  $k(s,t)$  in Eq.1. One algorithm ("reverse correlation") for obtaining this estimate is simply to take the average of the stimulus, as a joint function of spatial position and time lag, preceding each neuronal impulse (e.g., McLean and Palmer<sup>17</sup>) - see Fig.4B.

To the extent that a neuron really did behave like a truly linear system, this function would constitute a total description of its input-output behavior, and would allow a correct prediction of its response to any other stimulus. If the neuron behaves like a linear filter, but followed by an intensive nonlinearity (e.g., half-wave rectification, half-square operation), then the first kernel function will correctly represent the linear filter function.



**Fig.4** Measurement of 1st order kernel (spatiotemporal filter). **A.** Ternary white noise stimulus, consisting of 32 bars (only 8 shown here) at the neuron's optimal orientation, colored black, white or gray with equal probability. On each new exposure (5 or 10 ms, only 3 exposures shown), a fresh set of random grey levels are produced. Note key resolution parameters are exposure time and bar width. **B.** Occurrences of neuronal impulses (bottom trace) are correlated with preceding bar stimuli at each of a series of spatial positions (upper traces), as a function of the time lag between stimulus events and impulses.



**Fig.5** Measurement of 2nd order nonlinear interaction. Occurrences of neuronal impulses (bottom trace) within a defined time window are correlated with preceding pairs of bar stimuli at each of a series of bar positions (upper traces), as a function of their spatial ( $\Delta s$ ) and temporal ( $\Delta t$ ) separation.

The second order analysis is a similar cross-correlation of neuronal response with a record of the white noise, but now the correlation is made with *pairs* of preceding stimulus bars. Thus the complete second order analysis produces a kernel function of four variables, two spatial positions and two time lags (e.g., Marmarelis and Marmarelis<sup>25</sup>). However, such 5-dimensional data is difficult to visualize, and a collapsed "interaction function" is used here - it is calculated for spatial separations of  $\Delta s$  and temporal separations of  $\Delta t$  from one another, time-averaged over a temporal "window" slightly preceding each neuronal impulse (Fig.5). The average responses to the individual stimulus bars are subtracted out, to provide an estimate of the amount of nonlinear interaction between the pairs of stimuli. This calculation is also spatially averaged across all receptive field positions, to provide a function only of  $\Delta s$  and  $\Delta t$ , which will be referred to here as a "2nd order interaction function" (corresponding to the so-called "motion kernel" of Emerson et al<sup>27</sup>). Note that due to the superposition property, a purely linear system would produce zero for this function.

An instructive example to consider is the hypothetical Reichardt-type delayed-comparison model referred to earlier, and illustrated in Fig.6A: here it consists of photoreceptors separated by a distance  $D_0$ , whose signals are compared by an AND-gate. The left receptor signal is delayed by a time lag  $T_0$ ; thus the AND gate condition is met by a stimulus moving rightwards at a speed of  $D_0 / T_0$ . If we perform the ternary white noise analysis on this model, we obtain the first kernel function shown in Fig.6B, and the 2nd order interaction function in Fig.6C. Because the AND-gate is a (highly) nonlinear operator, the 2nd order analysis shows a strong nonlinear interaction at  $\Delta t = T_0$ ,  $\Delta s = D_0$ ; because of the space-averaging, there is also a mirror-symmetric interaction in the lower left quadrant. This example demonstrates that the method is well suited to measurement of the optimal spatial displacement ( $D_{opt}$ ) and optimal temporal separation ( $T_{opt}$ ) for 2-flash apparent motion.

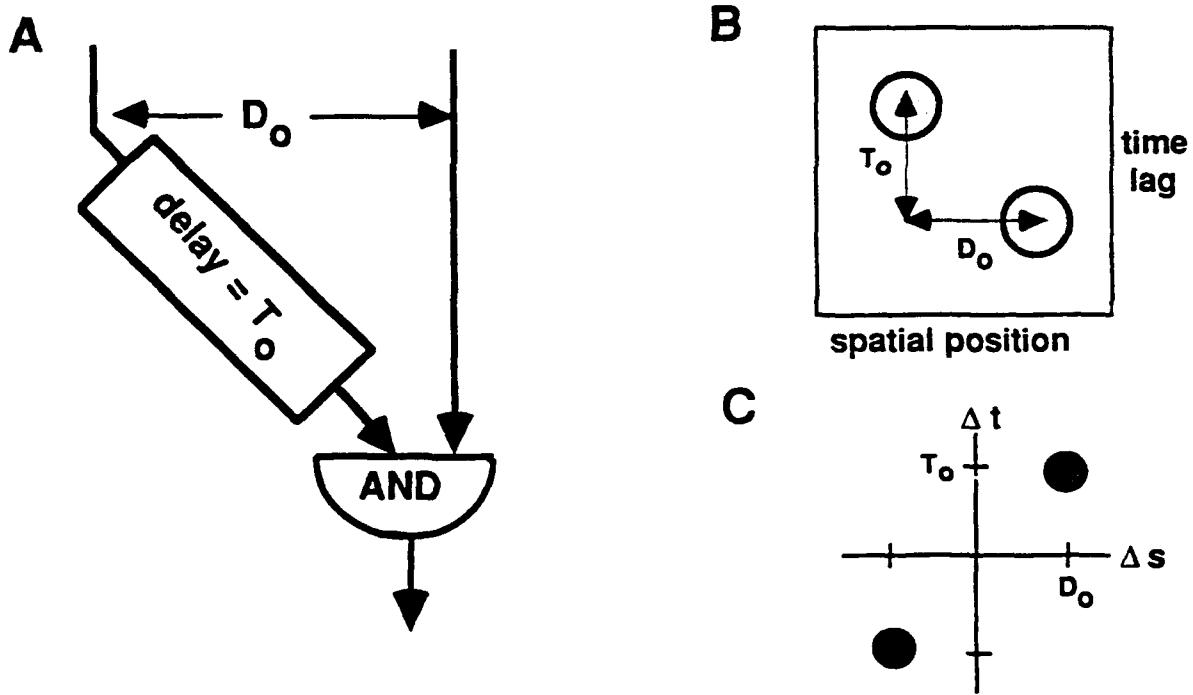


Fig.6 First and Second order white noise analysis for a delayed comparator model of direction selectivity. A. Model involving AND-gated comparison of inputs spatially separated by distance  $D_0$ , with the left one temporally delayed by  $T_0$ ; optimal stimulus consists of rightwards motion at speed of  $D_0 / T_0$ . B. First kernel function for this model. C. Second order interaction annulus for the same model.

Since this is a "pure nonlinear" system it might seem surprising to also observe a non-zero first kernel function (Fig.6B). But consider the model's requirement for producing a response, in relation to the ternary random stimulus: on each new exposure there is a 1/3 probability of any given bar being, e.g., white, independently for each exposure and thus for any time lag; so at any given time, there is a 1/9 probability of both input lines seeing a white bar. Thus the reverse

correlation algorithm will build up strong correlation measures at the two input positions, but with one of them delayed by  $T_O$  (Fig.6B). (For clarity, as well as to better mimic actual neuronal data, all signals in this model are subjected to a pure delay of about 50 ms). Note that for this particular model, the "single flash at a time" reverse-correlation method (McLean and Palmer<sup>17</sup>, DeAngelis et al<sup>11</sup>) would give a first kernel estimate of zero; those methods preclude estimation of 2nd order interactions, unless "double flashes" are used (Szulborski and Palmer<sup>28</sup>). This example also illustrates the possibility that a first-order spatiotemporal kernel which is correlated with a neuron's direction selectivity may be measured in a neuron having a quite nonlinear directional mechanism.

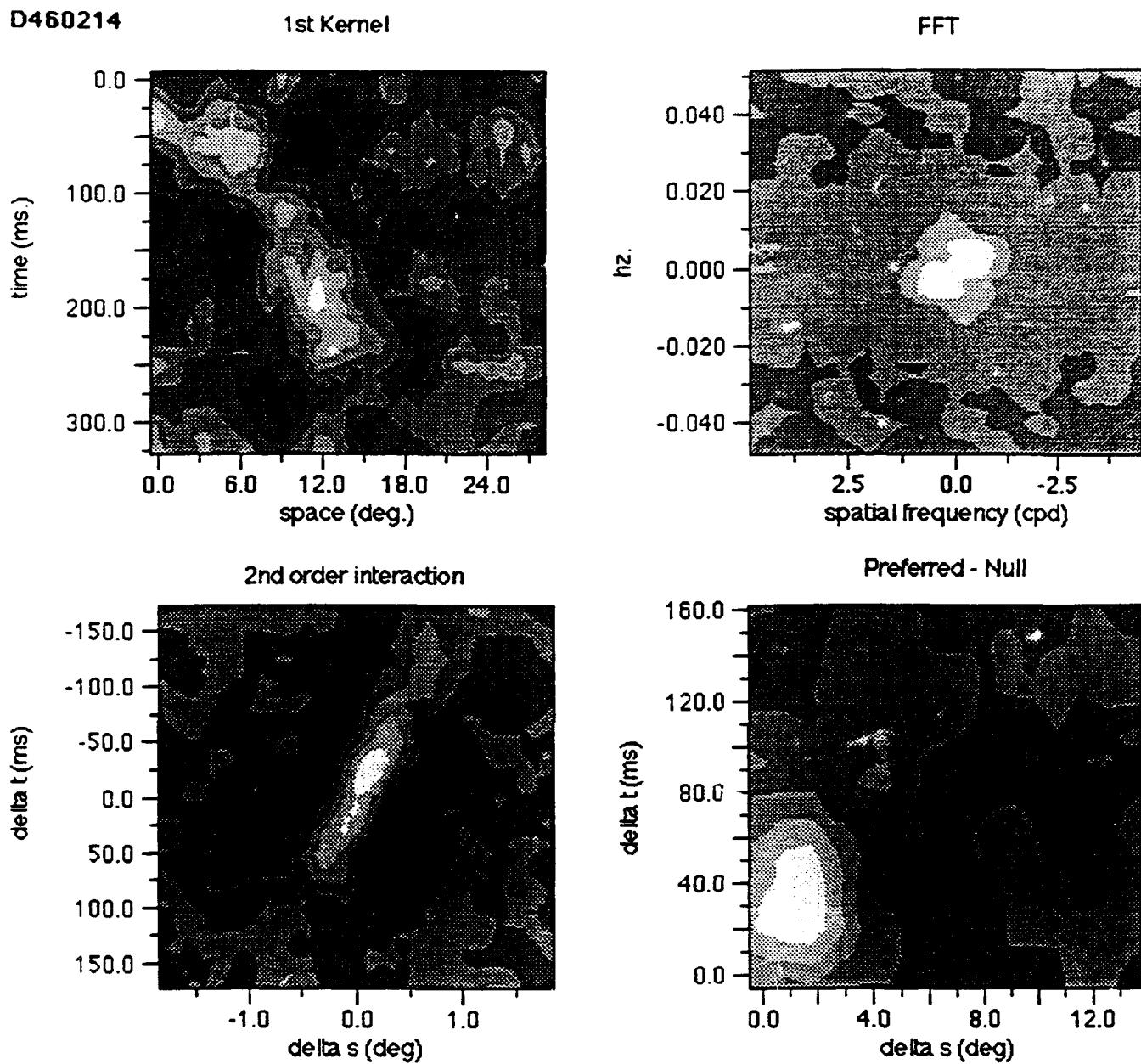
Fig.7 shows results of a ternary white noise measurement on a directionally selective Simple type cell. First notice beginning at a latency of about 50 msec, the spatially segregated positive (white) and negative (black) regions of the kernel function (Fig.7A), corresponding to the classical ON and OFF zones of the receptive field. At greater time lags (130-180 ms) these positions reverse in polarity, indicative of self-inhibitory or adaptive mechanisms. Interestingly, this reversal is much more pronounced for the OFF than for the ON zone, giving a crude kind of "space-time orientation" to the kernel function, somewhat like that of Fig.1D; and indeed this neuron was direction selective for stimuli moving rightwards relative to the plotted spatial coordinates.

Another way of visualizing this asymmetry is in the magnitude part of the Fourier transform of the kernel function, shown in Fig.7B. This plot shows a bandpass nature for both spatial and temporal frequency; the greater magnitude in the upper right / lower left quadrants, compared to the upper left / lower right, indicates a greater predicted response to rightward moving sinewave gratings, than for leftwards motion.

Fig.7C shows the 2nd order noniinear interaction plot - note that it also has a partially oriented nature, of the opposite direction to that of the first kernel (and therefore consistent with the sign of its directionality - compare Fig.6B-C). Though the plot is clearly oriented (non-separable), there are peaks in it corresponding to optimal spatial and temporal intervals. Thus the direction selectivity of this neuron has correlates in both the first and second order analyses.

Another useful kind of data reduction is to collapse the four quadrants of the 2nd order interaction into a plot of the net "preferred-minus-null" directionality as a function of the magnitude of  $\Delta s$  and  $\Delta t$ ; an example is seen in Fig.7D. This kind of plot has the advantage of always showing a separable dependence on  $\Delta s$  and  $\Delta t$ , thus allowing a well-defined estimate of  $D_{opt}$  and  $T_{opt}$  for neurons such as this one in which clear peaks are not so well-defined in the 2nd order interaction function (Emerson et al<sup>27</sup>).

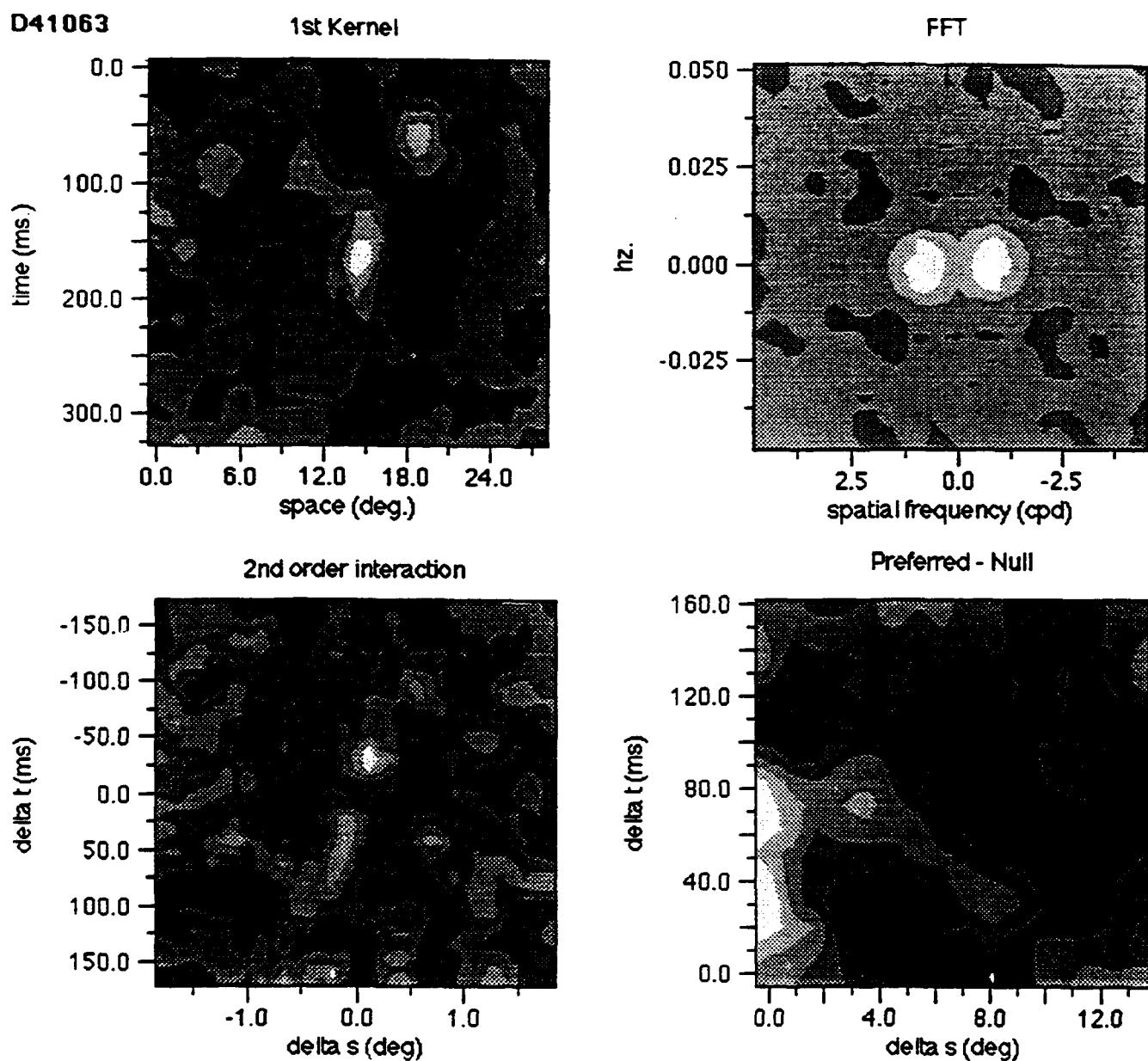
Note that in this example the  $D_{opt}$  value is about 1.2 deg. This value may be compared to an estimate of the neuron's spatial wavelength,  $\lambda$  (reciprocal of optimal spatial frequency), obtained from the spacing between ON and OFF regions in the first kernel (about 5 deg in Fig.7A, corresponding to a  $\lambda$  of about 10 deg. Thus the ratio of  $D_{opt} / \lambda$  is about 0.12 for this neuron; for most neurons examined so far, this ratio is generally somewhat less than 1/4 (quadrature phase, Eq. 2), in agreement with previous estimates based on discretely presented 2-flash stimuli (Baker and Cynader<sup>19</sup>) or multi-flash jumping gratings (Baker, Friend, and Boulton<sup>20</sup>). But unlike these previous studies, the white noise method affords the opportunity to also measure " $T_{opt}$ ", the optimal temporal pair-wise separation for direction-selectivity. Note in Fig.7D, the  $T_{opt}$  value is about 30 ms, which may be compared to an analogous "temporal wavelength" estimate made from the spacing of peaks and valleys in the temporal profile of the first kernel function, or alternatively from the first-order frequency response (Fig.7B). Again this neuron is typical of those examined so far, in that the  $T_{opt}$  value falls below a quadrature temporal phase relationship.



**Figure 7.** White noise analysis data from a Simple cell. A. First kernel function; at a temporal latency of about 50 ms, note positive (white) and negative (dark) zones, corresponding to discrete ON and OFF receptive field regions, with a weak diagonal "bridge", giving a space-time asymmetry (non-separability). B. Magnitude of Fourier transform of kernel function in A, plotted with origin at the center. Note the weak asymmetry, signifying direction selectivity. C. Second order interaction function, also plotted with origin at the center, showing oriented (non-separable) interaction. D. "Preferred-minus-Null" second order interaction, derived from C, showing a space-time separable directional interaction corresponding to (approximately) quadrature spatial and temporal phase, in comparison to the spacing of positive and negative regions in A.

A similar data analysis for a different direction-selective Simple cell is shown in Fig.8. In this case the first kernel (Fig.8A) shows adjacent ON and OFF regions, which both have very similar temporal profiles; thus the kernel is

nearly symmetrical ("space-time separable"), which can also be seen in the symmetry of the frequency response function in Fig.8B. Thus in this case the first order analysis fails completely to predict any direction-selectivity for the neuron.



**Figure 8.** White noise analysis data from another simple cell, in the same format as Fig.7. Note the lack of indication of direction selectivity in the first order analysis (A,B), and the approximate space-time separable, directional nonlinear interaction (C). In this example, the Preferred-minus-Null interaction plot (D) is less useful due to spatial resolution limits.

However, the 2nd order analysis (Fig.'s 8C) shows a clear directional interaction consistent with the neuron's preferred direction. Notice in this case that the interaction function is more nearly a separable function of  $\Delta s$  and  $\Delta t$ , with peaks to the upper right and lower left of the origin in Fig.8C, corresponding to a  $D_{opt}$  of about 0.2 deg and  $T_{opt}$  of about 25 ms. Interaction plots like Fig.8C are found in many cortical neurons, and are consistent with results published by

Emerson et al<sup>27</sup>, if one takes into account the spatio-temporal resolution (bar width and exposure time) relative to the spatial and temporal filtering of the neuron.

Data of this kind have been obtained for a variety of direction-selective Simple cells. In general a great diversity of results have been observed, indicating substantial heterogeneity across the population. A few neurons exhibit a clear spatiotemporally oriented the first kernel (like Fig.1D), while many have segregated positive and negative regions which are in an asymmetrical arrangement (like Fig.7A), and a few have little or no asymmetry (like Fig.8A); in general the degree of directionality predicted from the first order kernel is substantially less than that observed in response to drifting sinewave gratings. Many of these neurons show 2nd order interactions with asymmetries like those of Fig.7C-D or Fig.8C-D.

In the majority of cases, the direction of any asymmetry in the first or second order analyses is consistent with the cell's direction preference for drifting sinewave gratings. Neurons appear to vary greatly in the degree to which their direction-selectivity correlates with asymmetry of the first order kernel vs. that of the 2nd order interaction; many neurons show both.

With regard to space-time separability, a wide variety of results are observed. Fig.7 is not atypical, in that both first and second order plots show spatiotemporal "islands" that are "locally separable", connected by weaker oriented (non-separable) "bridges". In general the data from the population of neurons is largely consistent with both of the seemingly contradictory reports of Baker and Cynader<sup>29</sup> and of Emerson et al<sup>27</sup>, when one takes into consideration the small sample sizes of both studies, and the effectively lower spatiotemporal resolution used by the latter authors.

In most cases it is clear on qualitative grounds alone that the 2nd order interactions are not a trivial consequence of having a linear filter followed by a simple intensive nonlinearity. For example, the linear spatiotemporal filter corresponding to Fig.7A, followed by a half-square operator, would give a 2nd order interaction plot having  $D_{opt}$  and  $T_{opt}$  values corresponding to one-half, rather than one-quarter or less, of the spatial and temporal wavelengths. (Intuitively, the optimal 2-flash apparent motion stimulus to elicit a directional response from the filter of Fig.7A would be one whose space-time diagram can be translated so as to fall directly on the left ON region, then in the 2nd flash directly on the positive temporal reversal of the right OFF zone). Another example may be seen in Fig.8, in which the 2nd order interaction shows an asymmetry which does not follow from the structure of the first kernel.

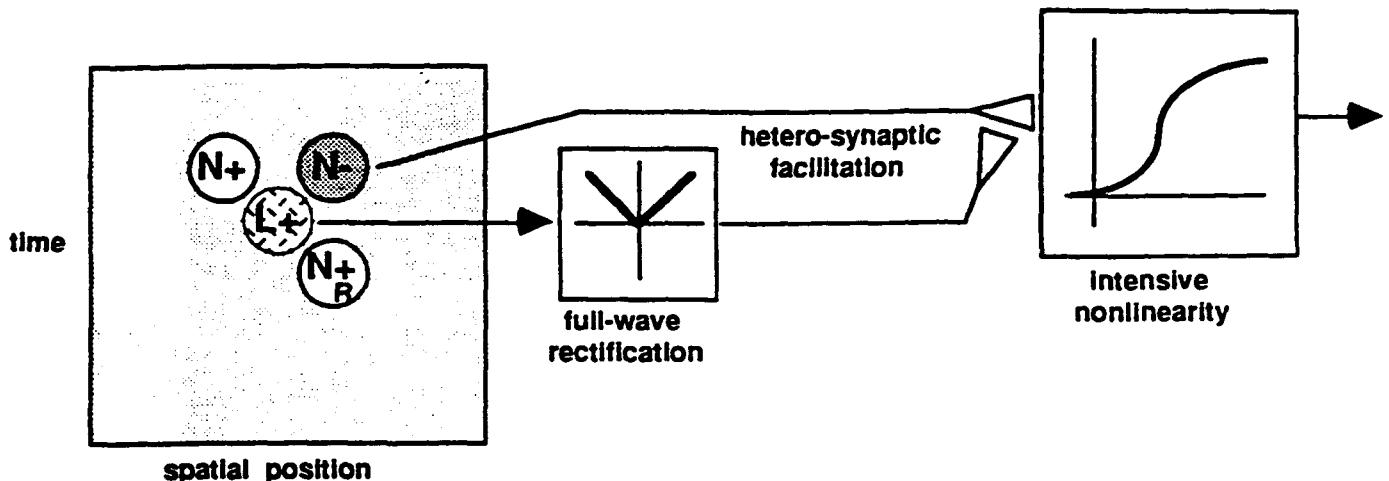
## Quasi-Linear Models

Thus in many cases, something a bit more nonlinear than a simple spatiotemporally linear filter followed by an intensive nonlinearity is needed to explain the data. Models aimed at accounting for such data will be termed "quasi-linear", because in spite of their nonlinearity they still exhibit many important correlates of their selectivity for direction of motion, spatial and temporal frequency, etc., in first-order data. The set of "quasi-linear" models should include "pure linear" models as a subset.

As a preliminary example, consider a first order kernel like that of Fig.7A, in which there is only a weak response at the spatiotemporal "quadrature point", yet the 2nd order analysis shows  $D_{opt}$  and  $T_{opt}$  values at quadrature spatial and temporal phase. In such a case it is reasonable to infer that a quadrature-phase input signal (e.g., from "lagged" type LGN afferents) might enter nonlinearly, in such a way as to have little influence on the first kernel. For example, the lagged input might be full-wave rectified before summing with the non-lagged inputs responsible for the measured first kernel; an additional possibility is that the non-lagged inputs might receive "heterosynaptic facilitation" (Nelson et al<sup>30</sup>) from the lagged inputs.

Such a "quadrature nonlinear" model is depicted in Figure 9. It consists of a relatively sustained ON zone on the left, and a spatially segregated, more transient OFF zone to the right; this will produce a first kernel having the spatio-temporal asymmetry of discrete positive and negative regions like those of Fig.7A, as well as a bandpass frequency response much as in Fig.7B. Because of the additional nonlinearity, the second order interaction will have a peak at approximately quadrature spatial and temporal phase, like that seen in the single unit measurements.

This model has some interesting similarities to the "STS" model of Marr and Ullman<sup>31</sup>, which used a nonlinear combination (AND-gating) of responses of two filters: a spatially odd-symmetric, sustained filter (itself formed by AND-gating of two adjacent odd-symmetric filters), and an even-symmetric, transient filter, the latter being half-wave rectified. Like the later, more linear models (Adelson and Bergen<sup>7</sup>; van Santen and Sperling<sup>6</sup>; Watson and Ahumada<sup>9</sup>), it involved combination of signals from filters that were in quadrature phase, both spatially (odd- and even-symmetric) and temporally (sustained and transient). This model differs in that the sustained (Lagged) filter is the one which enters in a more nonlinear way. The version shown in Fig.9 has only a single Lagged region for clarity, but at least in some neurons there are probably multiple, interdigitated Lagged and Nonlagged filters (Saul and Humphrey<sup>24</sup>), which would make both kinds of input have the same spatial-frequency selectivity.



**Figure 9.** An example of a "quasi-linear" model. Spatiotemporally linear signals from Lagged (L) and Non-lagged (N) inputs are combined nonlinearly. The nonlinearity may consist of full-wave (or half-wave) rectification of the lagged input, and/or a "hetero-synaptic facilitation" of the nonlagged signal by the lagged signal. In addition, there may be an intensive nonlinearity applied to the overall output. Neurons appear to vary considerably in the degree to which these different nonlinearities are present.

A more multiplicative or AND-gated type of interaction between lagged and non-lagged inputs would help sharpen direction selectivity (just as such an interaction between ON and OFF non-lagged inputs would sharpen selectivity for spatial frequency). This would incidentally contribute to a "half-square" appearance in measurements of the neuron's contrast response function (Albrecht and Giesler<sup>10</sup>).

Some of the range of diversity of simple cell receptive field organization might be accounted for in terms of variants of the above model. For example, if the relevant nonlinearity were a full-wave rectification of the Lagged input, it could produce a mixed ON-OFF zone, between clear ON and OFF regions, seen in some Simple cell receptive fields (e.g., Movshon, Thompson, and Tolhurst<sup>32</sup>). White noise data from a minority of neurons is clearly space-time oriented (much like Fig.1D), which could be produced with a linear combination rule for the Lagged input. The more common pattern of a weak oriented "bridge" between positive and negative regions, as in Fig.7, might be explained with a half-wave rather than a full-wave rectification of the Lagged input, or with a less purely nonlinear function for the combination of Lagged and Non-lagged inputs.

Another important, and probably wide-spread, variation of such a model, is based on having inhibition in the non-preferred direction of motion (e.g., like an "AND-NOT" operation - Barlow and Levick<sup>4</sup>) instead of facilitation in the preferred direction. Such a model would be a kind of "dual" of the above one, but the basic principles would be much the same; this would be consistent with reports that antagonists of the inhibitory neurotransmitter, GABA, can abolish the

direction selectivity of some cortical neurons (e.g., Wollman and Palmer<sup>33</sup>). The nonlinear combination rule might derive from hetero-synaptic inhibition, or from shunting inhibition.

Some of these nonlinearities of quasi-linear models might be merely the result of an imperfect realization of linear mechanisms from inherently nonlinear "wetware", while others might be functionally significant, e.g., for improving selectivity, or providing the competence to respond to some "non-Fourier" stimuli (see below).

The possibility of a diversity of mechanisms for direction selectivity should not come as a great surprise. It is very likely that direction selectivity (along with several other stimulus selective properties) is "learned" in early development (Chernenko and Cynader<sup>34</sup>), and so it should not be surprising if stimulus-dependent plastic mechanisms which strengthen connections of inputs that are activated in a correlated manner (Fregnac et al<sup>35</sup>) manage to build it from differing kinds of input signals in different neurons; this reasoning would apply also to the diversity of more seriously nonlinear mechanisms to be discussed below.

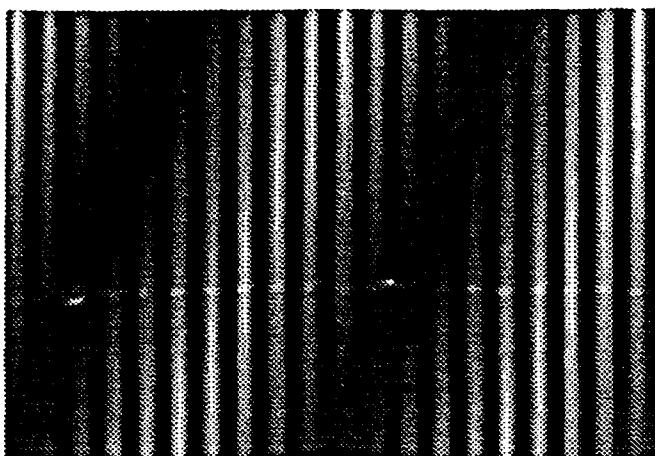
## Responses to "Non-Fourier" Stimuli

An alternative approach for uncovering behavior not predicted by linear models is to design stimuli to which they should not be direction selective, but whose direction of motion is readily discerned. Such stimuli have been termed "non-Fourier" (Chubb and Sperling<sup>36</sup>) because the spatiotemporal Fourier transform of the space-time diagram lacks any directional asymmetry. A possibly more intuitive way to think of this, is that the space-time diagram of the stimulus does not have consistently oriented regions of the same contrast sign. This approach is not only a way to detect nonlinear behavior, but may in some cases reveal mechanisms for detection of functionally significant stimuli (e.g., occlusion boundaries). The method has also been usefully applied to spatially two-dimensional stimuli, with the use of plaids (Movshon et al<sup>37</sup>) and illusory contours (von der Heydt and Peterhans<sup>38</sup>; Peterhans and von der Heydt<sup>39</sup>; Grossf and Shapley<sup>40</sup>). Here we will be more concerned with spatiotemporal examples, such as a traveling wave-front of contrast-reversal (Chubb and Sperling<sup>41</sup>) or a drifting contrast envelope (Albright et al<sup>42</sup>), in both cases with the "carrier" being spatially two-dimensional noise.

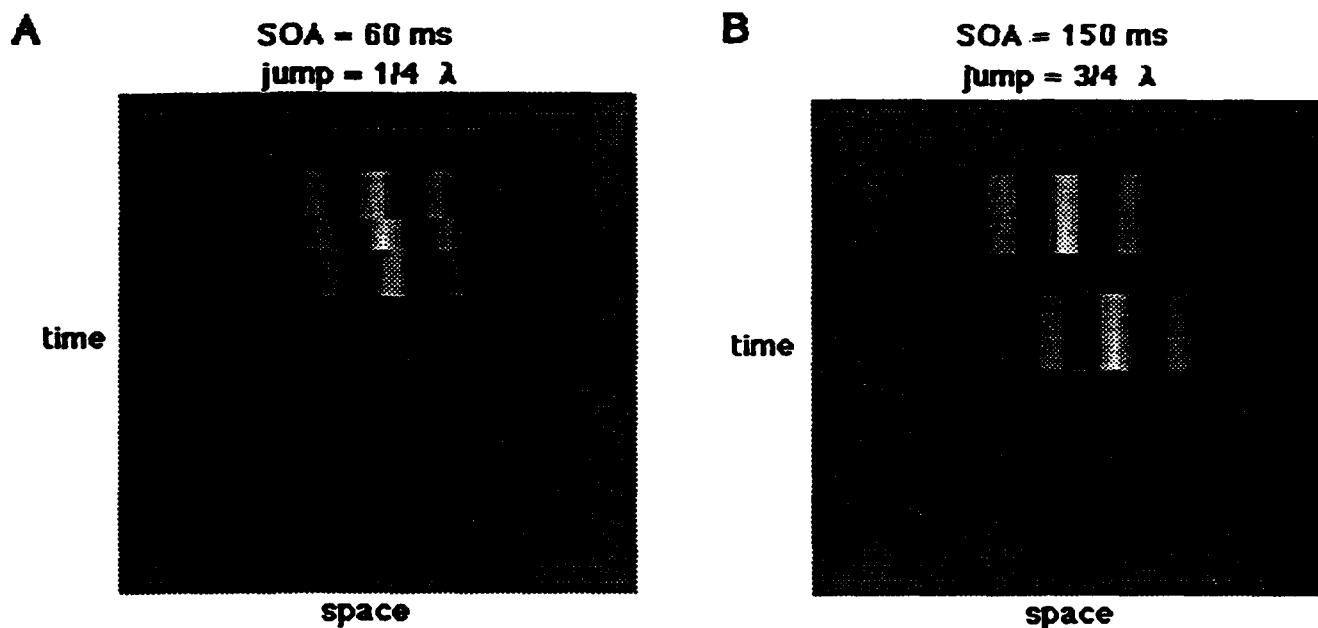
Recent work (Zhou and Baker<sup>43</sup>) has employed a drifting, low spatial frequency sinusoidal contrast envelope, which multiplies a stationary high spatial frequency sinewave carrier; this stimulus had previously been used in human psychophysical studies (e.g., Henning et al<sup>44</sup>). The space-time diagram of such a stimulus is shown in Fig.10. The carrier frequency is set well above the spatial frequency passband of the neuron; thus if the cell behaved linearly, it would not respond to this stimulus. However a substantial proportion of neurons in Areas 17 and 18 do respond, in a direction selective manner, but only for a surprisingly limited range of carrier frequencies. Some key points are that the spatial scale of the effective stimulus (carrier) is very different from the size and spacing of ON and OFF zones, and that the optimal envelope frequency is much lower than the neuron's optimal spatial frequency for conventional gratings; from the latter we may infer an integration of inputs outside the neuron's "classical" receptive field. Both these characteristics fall outside the scope of "quasi-linear" models.

Another example of cortical neuronal response to non-Fourier stimuli has followed on from a series of human psychophysics studies using "random Gabor kinematograms" (Boulton and Baker<sup>45</sup>; see accompanying paper in this volume<sup>46</sup>). Using a field of randomly placed Gabor micropatterns in two-flash apparent motion, we found that under some conditions (high density of micropatterns and short temporal interval), psychophysical performance was approximately predicted from the Fourier transform of the stimulus space-time diagram - "quasi-linear" behavior. However at lower densities or longer temporal intervals, this pattern of performance abruptly changed into one which was qualitatively different from that predicted by the stimulus power spectrum - "nonlinear" behavior. For the neurophysiological work a single Gabor function (a patch of sinewave grating, at the neuron's optimal spatial frequency, enclosed in a Gaussian envelope) was presented in two-flash apparent motion, at each of a series of spatial displacements and temporal intervals, at an optimal location in the neuron's receptive field. Fig.11 shows space-time diagrams for this stimulus for two sets of conditions. At a spatial displacement of one-fourth the wavelength of the Gabor sinusoid, corresponding to quadrature phase, and a relatively short temporal separation (stimulus onset asynchrony, SOA, of 60 ms), the Fourier transform shows clear directionality; notice that there is a consistently positive, oriented region corresponding to rightwards motion in the

space-time diagram (Fig.11A). On the other hand, Fig.11B shows the space-time diagram for the same stimulus, but with larger displacement and SOA values, for which the Fourier transform predicts no directionality.



**Figure 10.** Space-time diagram of leftwards moving "envelope stimulus" (a stationary, high spatial frequency sinewave multiplied by a moving, low spatial frequency sinewave). Ordinate is time, abscissa is spatial position.



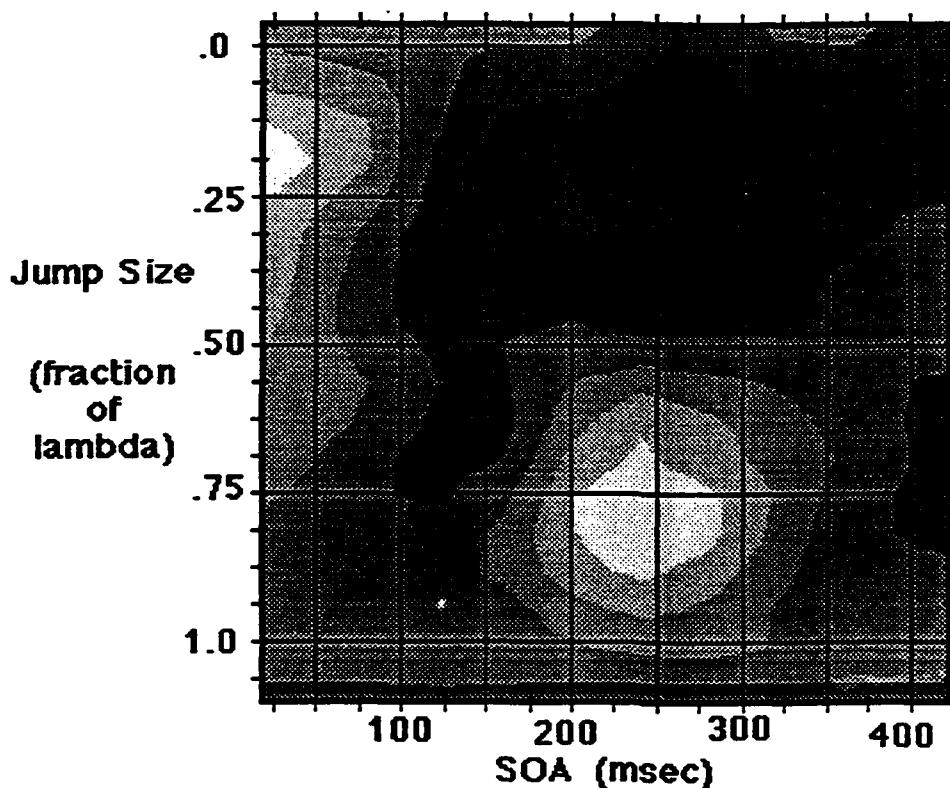
**Figure 11.** Space-time diagrams for two-flash Gabor stimulus. A. For a jump size of quadrature spatial phase and SOA of 60 ms, a linear model would predict good directionality. B. For a much larger jump size and SOA, the Fourier transform predicts no directionality.

Nearly all neurons tested in cat A17 and A18 (Boulton and Baker<sup>47</sup>) showed good direction selectivity for conditions like that of Fig.11A, corresponding to a "linear" prediction. In addition, however, many neurons also showed direction selectivity for some much larger values of displacement and/or SOA (Fig.11B). For example, Fig.12 shows the amount of direction selectivity (preferred-minus-null difference) as a intensity plot of displacement and SOA - notice the

"linear" behavior in the upper left part of the plot, and the nonlinear behavior in the lower right part, at a spatial displacement of about  $3/4 \lambda$  and an SOA of about 250 ms.

These very large displacements which can still elicit direction selectivity fall outside the "classical receptive field", and correspond to a coarser spatial scale than that which gives rise to the neuron's spatial frequency selectivity. Therefore this behavior would fall outside the "quasi-linear" range of mechanisms, and would be termed "nonlinear". Reinforcing this point still further, recent tests on a few neurons have demonstrated that the spatial frequency of the micropattern can be changed on the second flash, to be outside the neuron's spatial frequency passband, and still elicit direction selectivity. This is a neural correlate of similar findings in the human psychophysics experiments.

D1601.008



**Figure 12.** Amount of direction selectivity of a cortical neuron, plotted as a joint function of the spatial displacement (jump size) and temporal separation (stimulus onset asynchrony, SOA). Note the region of positive directionality in the upper left, corresponding approximately to a linear model prediction ("quasi-linear"), and the second positive region at a much larger jump size and SOA, which is not predicted by a linear model ("nonlinear").

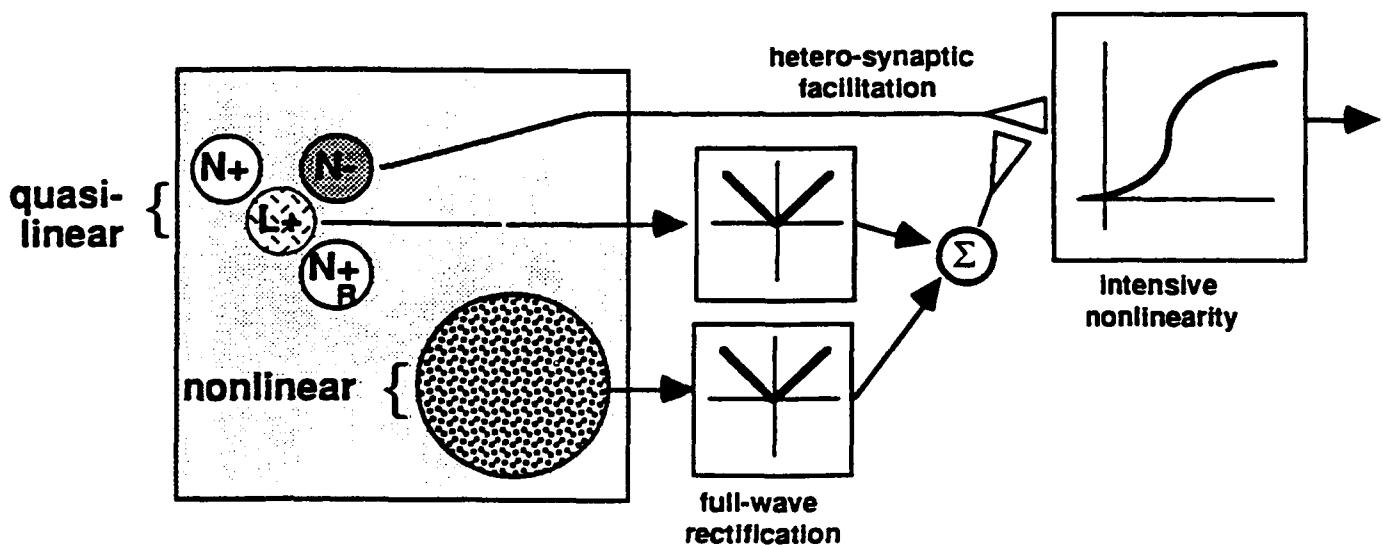
Certain common principles emerge from these examples of "nonlinear" behavior. Firstly, nonlinear behavior is always accompanied by quasi-linear behavior in the same neuron. While there certainly are cells that exhibit only quasi-linear behavior, we have yet to observe neurons in the visual cortex which are purely nonlinear in their behavior. Secondly, a given neuron has a consistent preferred direction of motion for both kinds of stimuli; this makes sense functionally, if the neuron really is a "labeled line" for the direction of motion, invariant with the nature of the moving stimulus (Albright<sup>42</sup>). Thirdly, all the "nonlinear" examples in this section involve stimuli that fall outside of the neuron's spatial

frequency / temporal frequency / orientation passband, and/or outside its classical receptive field. In one way or another, the effective stimuli are at a different spatial scale than that of the conventional sinewave grating spatial frequency tuning, and thus exhibit a more fundamental kind of nonlinearity than that of the "quasi-linear" models.

## A "Second Wave" Nonlinear Model

One such highly nonlinear model grew out of the work with two-flash apparent motion of Gabor micropatterns (Boulton and Baker<sup>46</sup>). Suppose that a briefly flashed Gabor function stimulus elicits a relatively short-latency, time-locked burst of activity in the visual cortex, which is quite specific for spatial frequency and orientation, as well as for retinal location. That is, only those neurons that are tuned to the Gabor's spatial frequency and orientation, and centered on its retinal location, will fire this early burst. But then there follows a longer-latency, weaker secondary wave of cortical activity, which spreads to adjacent cortical locations in a manner which is indiscriminate with respect to orientation or spatial frequency. Now assume that the motion correspondence mechanisms make use of whatever signals are available, whether they originate from the strong, punctate, specific first wave or from the weak, diffuse, non-specific second wave. Then the use of a high density of Gabor micropatterns and short SOA's would strongly favor motion correspondence based on the first wave; larger SOA's and/or lower densities would allow correspondence based on the second wave to manifest itself. Motion perception based on the second wave would be relatively weak in strength, would be invariant with changes in the orientation or spatial frequency from one frame to the next, and would show a  $D_{max}$  inversely dependent on density, since the lack of spatial frequency and orientation specificity would incur a serious vulnerability to the "correspondence problem" not suffered by the quasi-linear directionality based on the much more specific first wave (see Marr and Poggio<sup>48</sup> for a kind of quasi-linear model of stereopsis, in which the advantages of this specificity for solving the correspondence problem are discussed). For example, such a mechanism would be good at detecting the motion of a rotating soccer ball, but only if it does not appear against a background of other moving objects.

An example of how such a "second wave" model might be implemented is shown in Fig.13. It has been cast into a form to be comparable to the example "quasi-linear" model of Fig.9. The hashed region in the space-time filter function, labeled "nonlinear", is meant to denote input from the second wave signal. Obviously there are a variety of alternative, but similar, ways in which these signals might be combined to form the direction-selective response of a single neuron.



**Figure 13.** An example of a nonlinear, "2nd Wave" model, cast in a similar format to the quasi-linear example in Fig.9. The textured region in the lower right of the spatiotemporal filter represents a delayed input from many signals which are not specific for the spatial frequency or orientation of the quasi-linear filter.

The origin of the second wave signal might be cortical (e.g., a sum of Complex cells, tuned to many spatial frequencies and orientations), or subcortical (e.g., a population of LGN afferents, which are not organized into spatially quasi-linear filters). In the primate visual system, there is evidence that the first wave signal is based on the magnocellular pathway, while the second wave signal uses the parvocellular pathway - see Boulton and Baker<sup>46</sup>.

It should be clear that this is not a viable model for other kinds of nonlinear behavior - for example, it would not work in this form to explain neuronal responses to envelope stimuli (see Zhou and Baker<sup>43</sup>), primarily due to the narrowband tuning to the carrier frequency observed in those experiments. Again, it seems likely that there is a diversity of kinds of nonlinearity, which enable response to a much wider range of possible stimuli than would be visible to a single generic kind of motion detector.

## Discussion and Conclusions

The discovery of nonlinearities of visual processing should not be taken as an incitement to totally discard previous concepts based on Fourier analysis and spatial frequency "channels"; instead it appears more appropriate to build on these ideas with supplementary modifications (quasi-linear) or parallel sub-systems (nonlinear). For many purposes it seems likely that linear systems approaches will continue to prove useful for understanding of some aspects of vision, while other, more nonlinear models will be needed to account for other phenomena, perhaps like the use of "wave" vs. "particle" concepts of light, in the domain of physics.

It is conceivable that not all responses to "non-Fourier" stimuli are mediated by "nonlinear" mechanisms, as the term is used here. That is, in some cases a "quasi-linear" mechanism might mediate response to a "non-Fourier" stimulus. An example may be seen in Chubb and Sperling's<sup>41</sup> model involving nonlinear combination of sustained and transient signals; this model has some interesting similarities to that of Fig.9. The key distinction being made here is in terms of whether the processing is at the *same spatial scale* as the neuron's (first order) spatial properties.

Note that white noise data generally do not show signs of "non-linear" mechanisms: the analyses reveal structure only at a level corresponding to the spatiotemporal frequency passband of the cell, and nothing outside of the classical receptive field. Thus it seems that the white noise method is good for the analysis of nonlinearities of quasi-linear systems, but not of the brutally nonlinear ones. Some possible reasons for this are the following: (1) non-linear responses are often relatively weak, and so their signatures in the cross-correlations might be too small to show up in against the often noisy background; (2) in some cases (envelope-responsiveness) the bar widths which can effectively drive the cell are at too coarse a spatial scale for the relevant nonlinearity (carrier spatial frequency tuning); (3) some nonlinearities might be 4th order (e.g., a sum of some parallel, full-wave rectified signals, which is again full-wave rectified), and not have correlates in lower (first and second) order analyses.

## ACKNOWLEDGEMENTS

We thank Michael Moskovich and Ken Charles for computer programming assistance, and Sean Friend and Yi-Xiong Zhou who assisted with some experiments. We are grateful to Rhone-Poulenc Pharma for their generous donation of Gallamine Triethiodide. This research was supported by Canadian MRC (MA 9685) and Stairs Memorial Fund grants to C.B., and a Human Frontiers Postdoctoral Fellowship to J.B.

## REFERENCES

<sup>1</sup>Saul, A.B., and Humphrey, A.L., "Evidence of input from lagged cells in the lateral geniculate nucleus to simple cells in cortical area 17 of the cat", *J.Neurophysiol.* 68: 1190-1208 (1992).

<sup>2</sup>Eysel, U.T., Muche, T., and Worgötter, F., "Lateral interactions at direction-selective striate neurones in the cat demonstrated by local cortical inactivation", *J.Physiol.* 399: 657-675 (1988).

<sup>3</sup>Goodwin, A.W., Henry, G.H., and Bishop, P.O., "Direction selectivity of simple striate cells: properties and mechanism", *J. Neurophysiol.* 38: 1500-1523 (1975).

<sup>4</sup>Barlow, H.B., and Levick, W.R.(Jr.), "The mechanism of directionally selective units in the rabbit's retina", *J. Physiol. Lond.* 178: 477-504 (1965).

<sup>5</sup>Reichardt, W., "Nervous integration in the facet eye", *Biophys.J.* 2: 121-143 (1962).

<sup>6</sup>van Santen, J.P.H. and Sperling, G., "Elaborated Reichardt detectors", *J.Opt.Soc.Am.* 2: 300-321 (1985).

<sup>7</sup>Adelson, E.H. and Bergen, J.R., "Spatiotemporal energy models for the perception of motion", *J.Opt.Soc.Am.* 2: 284-299 (1985).

<sup>8</sup>Heeger, D.J., "Nonlinear model of neural responses in cat visual cortex", pp.119-133, in *Computational Models of Visual Processing*, Ed. M.S. Landy and J.A. Movshon, MIT Press: Cambridge,MA 1991.

<sup>9</sup>Watson, A.B. and Ahumada, A.J.(Jr.), "Model of human visual-motion sensing", *J.Opt.Soc.Am.* 2: 322-342 (1985).

<sup>10</sup>Albrecht, D.G., and Giesler, W.S., "Motion selectivity and the contrast-response function of simple cells in the visual cortex", *Vis.Neurosci.* 7: 531-546 (1991).

<sup>11</sup>DeAngelis, G.C., Ohzawa, I., and Freeman, R.D., "Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation", *J.Neurophysiol.* (1993, in press).

<sup>12</sup>Tolhurst, D.J. and Movshon, J.A., "Spatial and temporal contrast sensitivity of striate cortical neurones", *Nature Lond.* 257: 674-675 (1975).

<sup>13</sup>Friend, S.M., and Baker, C.L.(Jr.) "Spatio-temporal frequency separability in area 18 neurons of the cat", *Vision Res.* (in press, 1993).

<sup>14</sup>McLean, J., and Palmer, L.A., "Responses of simple cells in areas 17 and 18 of the cat in the spatiotemporal frequency domain", *Invest.Ophthal.Vis.Sci.* 30: 111 (1989).

<sup>15</sup>Reid, R.C., Soodak, R.E., and Shapley, R.M., "Directional selectivity and the spatiotemporal structure of the receptive fields of simple cells in cat striate cortex", *J.Neurophysiol.* 66: 505-529 (1991).

<sup>16</sup>Tolhurst, D.J., and Dean, A.F., "Evaluation of a linear model of directional selectivity in simple cells of the cat's visual cortex", *Vis.Neurosci.* 6: 421-428 (1991).

<sup>17</sup>McLean, J. and Palmer, L.A., "Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat", *Vision Res.* 29: 675-679 (1989).

<sup>18</sup>Baker, C.L.(Jr.), "Spatial and temporal frequency selectivity as a basis for velocity preference in cat striate cortex neurons", *Visual Neuroscience* 4: 101-113 (1990).

<sup>19</sup>Baker, C.L.(Jr.) and Cynader, M.S., "Spatial receptive field properties of direction selective neurons in cat striate cortex". *J.Neurophysiol.* 6: 1136-1152 (1986).

<sup>20</sup>Baker, C.L.(Jr.), Friend, S.M. and Boulton, J.C., "Optimal spatial displacement for direction selectivity in cat visual cortex neurons", *Vision Res.* 31: 1659-1668 (1991).

<sup>21</sup>Baker, C.L.(Jr.), Baydala, A., and Zeitouni, N., "Optimal displacement in apparent motion", *Vision Res.* 29: 849-859 (1989).

<sup>22</sup>Boulton, J.C., and Baker, C.L.(Jr.), "Different parameters control motion perception above and below a critical density", *Vision Res.* (1993, in press).

<sup>23</sup>Mastronarde, D.N., "Two classes of single-input X-cells in cat lateral geniculate nucleus. I. Receptive-field properties and classification of cells", *J. Neurophysiol.* 57: 357-413 (1987).

<sup>24</sup>Saul, A.B., and Humphrey, A.L., "Evidence of input from lagged cells in the lateral geniculate nucleus to simple cells in cortical area 17 of the cat", *J.Neurophysiol.* 68: 1190-1208 (1992).

<sup>25</sup>Marmarelis, P.Z., and Marmarelis, V.Z., *Analysis of physiological systems: The white noise approach*. New York: Plenum Press (1978).

<sup>26</sup>Suttorp, E., "A deterministic approach to nonlinear systems analysis", pp 171-220, in *Nonlinear Vision: Determination of Neural Receptive Fields, Function, and Networks*, Ed. R.B. Pinter and B. Nabat, CRC Press: Boca Raton, 1992.

<sup>27</sup>Emerson, R.C., Citron, M.C., Vaughn, W.J., and Klein, S.A., "Nonlinear directionally selective subunits in complex cells of cat striate cortex", *J. Neurophysiol.* 58: 33-65 (1987).

<sup>28</sup>Szulborski, R.G., and Palmer, L.A., "The two-dimensional spatial structure of nonlinear subunits in the receptive fields of complex cells", *Vision Res.* 30: 249-254 (1990).

<sup>29</sup>Baker, C.L.(Jr.) and Cynader,M.S., "Space-time separability of direction selectivity in cat striate cortex neurons", *Vision Res.* 28: 239-246 (1988).

<sup>30</sup>Nelson, S.B., Toth, L.J., and Sur, M., "Spatial and temporal integration of synaptic inputs by visual cortex neurons", *Soc.Neurosci.Abst.* 18: 99.4 (1992).

<sup>31</sup>Marr, D. and Ullman, S., "Directional selectivity and its use in early visual processing", *Proc.R.Soc.Lond. B* 211: 151-180 (1981).

<sup>32</sup>Movshon, J.A., Thompson, D.D., and Tolhurst, D.J., "Spatial summation in the receptive fields of simple cells in the cat's striate cortex", *J. Physiol. Lond.* 283: 53-77 (1978).

<sup>33</sup>Wollman, D.E., and Palmer, L.A., "The effects of GABA blockade on the spatiotemporal receptive field structure of neurons in cat striate cortex", *Invest.Ophthal.Vis.Sci.* 34: 1028 (1993).

<sup>34</sup>Cynader, M.S., and Chernenko, G., "Abolition of direction selectivity in the visual cortex of the cat", *Science* 193: 504-505 (1976).

<sup>35</sup>Fregnac, Y., Shultz, D., Thorpe, S., and Bienenstock, E., "A cellular analogue of visual cortical plasticity", *Nature* 333: 367-370 (1988).

<sup>36</sup>Chubb, C., and Sperling, G., "Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception", *J.Opt.Soc.Am.A* 5: 1986-2007 (1988).

<sup>37</sup>Movshon, J.A., Adelson, E.A., Gizzi, M.S., and Newsome, W.T., "The analysis of moving visual patterns. In: Pattern Recognition Mechanisms. Ed. C.Chagas, R. Gattass, and C. Gross. Rome: Vatican Press, Pont.Acad.Sci.Scr.Varia 54: 117-151 (1985).

<sup>38</sup>von der Heydt, R., and Peterhans, E., "Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity", *J.Neurosci.* 9: 1731-1748 (1989).

<sup>39</sup>Peterhans, E., and von der Heydt, R., "Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps", *J.Neurosci.* 9: 1749-1763(1989).

<sup>40</sup>Grosof, D.H., Shapley, R.M., and Hawkin, M.J., "Monkey striate responses to anomalous contours?", *Invest.Ophthal.Vis.Sci.(Suppl)* 33: 2822 (1992).

<sup>41</sup>Chubb, C., and Sperling, G., "Two motion perception mechanisms revealed through distance-driven reversal of apparent motion", *Proc.Natl.Acad.Sci.* 86: 2985-2989 (1989).

<sup>42</sup>Albright, T.D., "Form-cue invariant motion processing in primate visual cortex", *Science* 255: 1141-1143 (1992).

<sup>43</sup>Zhou, Y-I., and Baker, C.L.(Jr.), "A processing stream in mammalian visual cortex neurons for non-Fourier responses", *Science* (1993, in press).

<sup>44</sup>Henning, G.B., Hertz, B.G., and Broadbent, D.E., "Some experiments bearing on the hypothesis that the visual system analyses spatial patterns in independent bands of spatial frequency", *Vision Res.* 15: 887-897 (1975).

<sup>45</sup>Boulton, J.C., and Baker, C.L.(Jr.), "Dependence on stimulus onset asynchrony in apparent motion: Evidence for two mechanisms", *Vision Res.* (1993, in press).

<sup>46</sup>Boulton, J.C., and Baker, C.L.(Jr.), "Psychophysical evidence for both a 'quasi-linear' and a 'nonlinear' mechanism for the detection of motion", *Proc. S.P.I.E. (this issue)*.

<sup>47</sup>Boulton, J.C., and Baker, C.L.(Jr.), "Nonlinear responses to apparent motion in cat visual cortex neuron", *Soc.Neurosci. Abstr.* 17: 404.9 (1991).

<sup>48</sup>Marr, D., and Poggio, T., "A computational theory of human stereo vision", *Proc.R.Soc.Lond.Ser.B.* 204: 301-328 (1979).

# Psychophysical evidence for both a "quasi-linear" and a "nonlinear" mechanism for the detection of motion.

Jane C. Boulton and Curtis L. Baker Jr.

J.B. Utrecht Biophysics Research Institute, Princeton Plein 5, Post Bus 80 000, 3508 TA Utrecht, The Netherlands.  
C.B. McGill Vision Research Unit, 687, Pine Avenue West, Rm H4-14, Montreal, Quebec, Canada H3A 1A1.

## ABSTRACT

A "random Gabor Kinematogram" stimulus provides the opportunity to demonstrate Fourier and non-Fourier motion perception, and discontinuities of performance from one to the other, in a way which supports the existence of categorically distinct underlying mechanisms.

Two frame apparent motion was used with a stimulus comprised of micro-patterns randomly distributed across the visual field. The micro-patterns were Gabor functions which contain a narrow band of spatial frequencies and orientations whilst maintaining a local nature in space. Psychophysical techniques were used to assess the detection of motion of this stimulus; two underlying processes were identified and characterized. For short temporal intervals and spatially dense stimuli, the response of the visual system can be predicted from the direction information in the spatio-temporal Fourier power spectrum of the stimulus: a "quasi-linear" mechanism. For longer temporal intervals and spatially sparse stimuli, detection of motion is NOT predictable from the information in the spatio-temporal Fourier power spectrum. Performance is independent of the spatial frequency content and orientation of the micro-patterns, but is limited by the "density" of stimulus elements along the axis of motion: a "nonlinear" mechanism.

It is proposed that the "nonlinear" mechanism is mediated by the parvocellular retina-cortical pathway, and the "quasi-linear" by the magnocellular pathway.

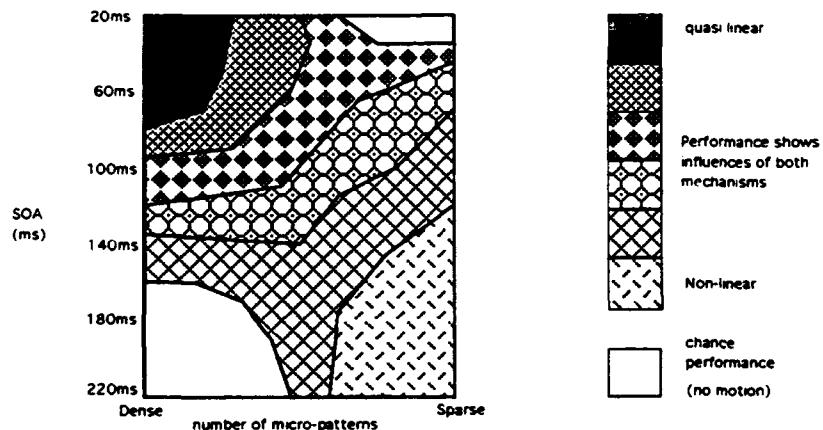
## 1. INTRODUCTION.

There are several reports in the literature suggesting two processes for the detection of motion with differential spatial and temporal sensitivities<sup>1,2,3,4,5</sup>. A proposal by Braddick<sup>1</sup> was based on the spatial displacement over which motion could be detected<sup>2,3,6,7,8</sup>. The "short range" process was thought to reflect properties of low level vision and to operate over small spatial and temporal displacements. The "long range" process was proposed to mediate apparent motion for larger displacements and longer time intervals; it was suggested to be characteristic of higher level processes and to be exemplified by the classical "phi" motion<sup>9</sup>. Recently the idea of two mechanisms serving motion perception has been phrased in terms of "first order" and "second order" motion processes<sup>4</sup>. The "first order" process reflects properties of low level vision which can be modelled by a system with early linear filters; that is, the ability to detect motion is predictable from the spatial-temporal Fourier power spectrum of the stimulus. Models of this type include Elaborated Reichardt Detectors (ERD)<sup>10,11,12,13</sup>. The "second order" process is proposed to mediate the perception of motion in a stimulus which has no overall directional component in the Fourier domain<sup>4</sup>.

A criticism of both these previously proposed dichotomies is that they reflect differences in the choice of stimuli, rather than qualitative differences in the underlying motion detection mechanisms<sup>14,15</sup>. With the use of random Gabor kinematograms, we have been able to use a single stimulus to reveal two qualitatively different processes for the detection of motion. These two processes have similar characteristics to the "first order" and "second order" processes proposed by Chubb and Sperling<sup>4</sup>. We constructed a stimulus in which many identical micro-patterns are randomly positioned throughout the stimulus field. These micro-patterns are Gabor functions which are narrow-band in both spatial frequency and orientation, in keeping with evidence that the visual system processes information via orientated, spatial frequency selective channels<sup>16,17,18,19,20,21</sup>. This form of stimulus construction also provides the opportunity to independently manipulate local stimulus attributes such as the *size* (spatial extent), *content* (spatial frequency) and *density* of the stimulus features (micro-patterns).

We have used this stimulus (a random Gabor kinematogram) to reveal two mechanisms. The stimulus was presented in two flash apparent motion with a variable temporal offset between the beginning of the first and second flashes (stimulus onset asynchrony, SOA), the number of micro-patterns in the stimulus was also varied. The two mechanisms were found to have different sensitivities to these stimulus parameters which are summarized in the schematic diagram shown in figure 1. The black area represents the spatial temporal parameters under which direction discrimination performance is almost perfectly predicted from the direction information in the spatio-temporal Fourier power spectrum. The "quasi-linear" mechanism is therefore, most sensitive for stimuli dense in both space and time. The textured areas show how performance changes from the linear prediction to the other extreme, where for stimuli sparse in both space and time, the "nonlinear" mechanism is most sensitive. Here performance is unrelated to the direction information in the spatio-temporal Fourier power spectrum. The maximum displacement for which the correct direction of motion is perceived is dependent on the number of micro-patterns in the stimulus. In this article we present evidence to support the hypothesis that the two mechanisms are qualitatively different and possibly mediated via separate retina-cortical pathways.

Schematic diagram of the sensitivity of the two mechanisms



## 2. METHODS.

For a detailed description of the methods and apparatus, see Boulton and Baker<sup>22,23,24</sup>. The stimuli consisted of micro-patterns distributed semi-randomly across the visual field. The micro-pattern was a Gabor function, that is a one-dimensional sinewave grating multiplied by a two-dimensional Gaussian window:

$$L(x,y) = L_0 \{1 + C \exp[-(x^2/2\sigma_x^2 + y^2/2\sigma_y^2)] \cdot \cos(2\pi x/\lambda + \phi)\} \quad (1)$$

where  $L_0$  = mean luminance;  $C$  = contrast;  $\sigma_x$  = horizontal Gaussian width parameter;  $\sigma_y$  = vertical Gaussian width parameter;  $\lambda$  = wavelength of the cosine wave;  $\phi$  = phase of cosine wave.

The micro-patterns were placed in two strips across the top and bottom of the stimulus field so as to confine the stimulus in eccentricity (about 4 degrees vertically) and to prevent the observers from paying attention to a fortuitous stimulus "feature" (e.g. a relatively isolated micro-pattern) close to the fixation mark. On each trial, micro-patterns were placed on a notional grid of  $X$  columns and 3 rows in each strip; each micro-pattern position was randomly "jittered" by 1/3 of the grid spacing about the grid location, to prevent a periodicity effect. The number of columns ( $X$ ) was dependent on the experimental condition, the stimulus density equals the number of micro-patterns per stimulus row, which equals  $X$ . Values of positional jitter were independently selected each trial for each micro-pattern. The stimulus field of randomly distributed micro-patterns was presented in one position for 100ms then displaced by a specific number of pixels to either the left or the right (with wrap-around at the display boundaries), and presented for another 100ms in the new position. Whenever there was no stimulus present, (including the inter-stimulus interval when the SOA exceeded 100ms) the stimulus field was

of mean luminance,  $L_0$ . If the two stimulus presentations overlapped (i.e. the SOA was less than 100ms) then the modulation of the two stimuli around the mean luminance were linearly summed. Unless otherwise specified the contrast of the micro-patterns was 14dB (20%). Observers maintained gaze on a centrally located fixation point, and initiated each trial via a button press. A method of constant stimuli was used for a range of displacements, with a two alternative forced choice procedure for the discrimination of the direction of motion. Psychometric functions of percent errors in direction discrimination were collected as a function of the displacement of the stimulus.

### 3. RESULTS

#### 3.1 Typical results

The "quasi-linear" mechanism. When the stimulus is comprised of many micro-patterns and presented in two flash apparent motion with a short SOA ( $<= 100$ ms), direction discrimination performance can be predicted from linear systems theory. That is, performance is proportional to the direction information in the spatio-temporal Fourier power spectrum of the stimulus. An example of direction discrimination performance as a function of the displacement of the micro-patterns is shown in figure 2a. The perceived direction of motion is related to the periodicity of the cosine component within the micro-patterns. When the micro-patterns are displaced by 3/4 of this periodic cycle reversed motion is perceived (for a detailed discussion of these results see Boulton and Baker<sup>23,24</sup>).

The "nonlinear" mechanism. As the number of the micro-patterns in the stimulus is reduced and/or the SOA increased, performance displays a sharp transition after which it is no longer related to the periodicity within the micro-patterns. For sparsely populated stimuli and long SOAs performance is unrelated to the direction information in the Fourier power spectrum. The maximum displacement for which direction is correctly perceived is inversely dependent on the number of micro-patterns along the axis of motion irrespective of their contents. An example of the performance of this "nonlinear" mechanism is shown in figure 2b. Note that for a displacement of 0.75  $\lambda$  performance is near perfect where as in figure 2a, this displacement produced 90% errors in direction discrimination, indicating that the "wrong" direction was perceived (for a detailed discussion of these results see Boulton and Baker<sup>23,24</sup>).

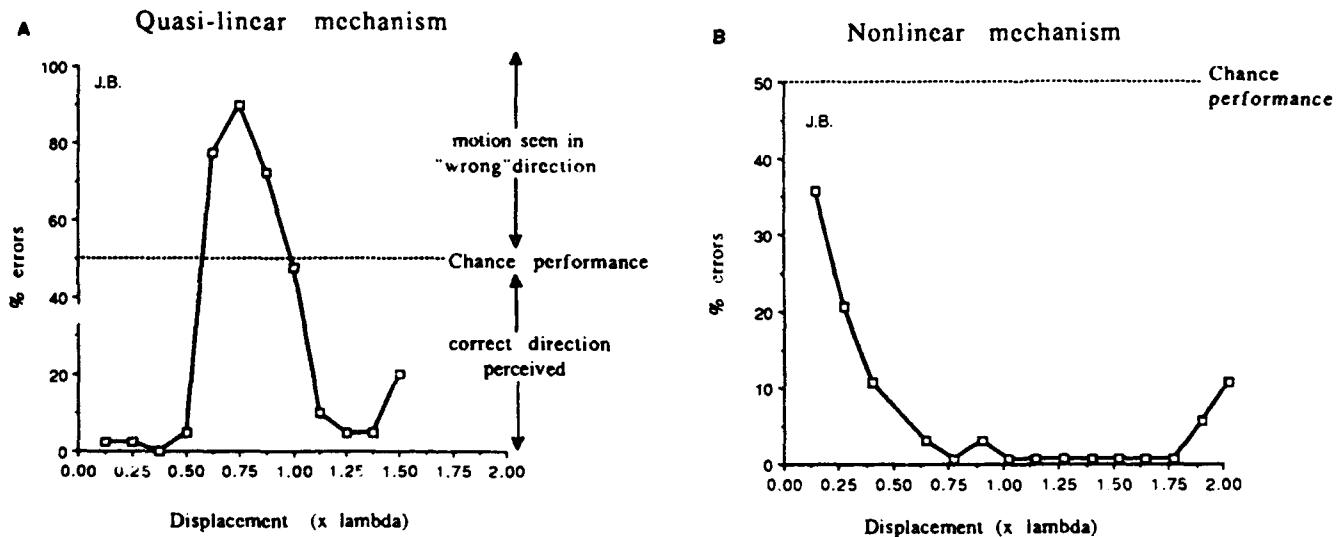


Figure 2

(a) The micro-patterns for both the first and the second exposure had a central spatial frequency of 2.25c/d ( $\lambda = 16$  pixels) and  $\sigma_x = \sigma_y = 0.75\lambda$ . The SOA = 60ms (the two exposures overlap for 40ms), and the stimulus field was densely packed with micro-patterns (11 per row). Percentage errors in direction discrimination are plotted as a function of the displacement (multiples of  $\lambda$ ). (b) The micro-patterns were the same as in (a), but the SOA was increased to 140ms (40ms ISI) and the number of micro-patterns per row was reduced to 5 per row.

In the above experiments the micro-patterns in the stimulus have remained unchanged throughout the motion sequence.

In the following two sections we manipulate the contents of the micro-patterns between the two exposures that constitute the motion sequence.

### 3.2 Spatial frequency selectivity

The "quasi-linear" mechanism. The number of micro-patterns in the stimulus field, and the temporal offset between the two exposures of the motion sequence (SOA) were selected for optimal performance of the "quasi-linear" mechanism. That is, an SOA of 60ms, and a dense pattern. In the first exposure of the motion sequence, the spatial frequency of the micro-patterns was 2.25c/d. In the second exposure the micro-patterns were of the same size (spatial extent) as in the first exposure but the spatial frequency differed. Performance for direction discrimination is shown as a function of the displacement of the micro-patterns in figure 3a. The continuous line shows the results displayed in figure 2a, i.e. no change in spatial frequency between the exposures of the motion sequence. The two dashed lines show results for when the spatial frequency of the second exposure differed by  $\pm$  one octave from the spatial frequency in the first exposure. In these cases performance is reduced to chance (50% errors) demonstrating that the "quasi-linear" mechanism is sharply tuned to spatial frequency. The "quasi-linear" mechanism does not establish correspondence between spatial patterns differing by one octave in frequency.

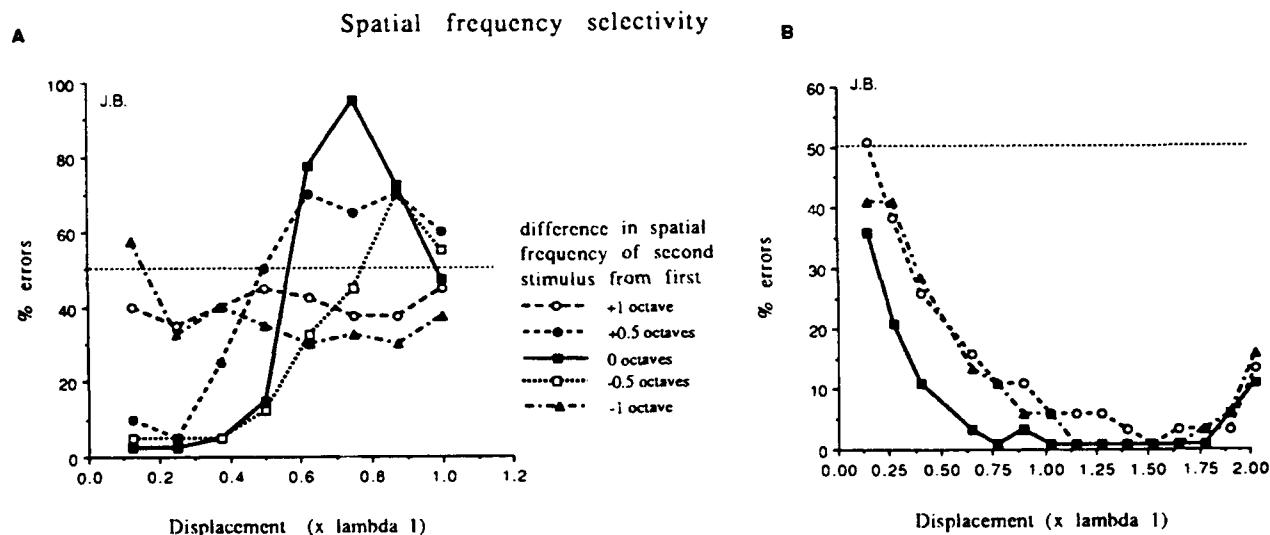


Figure 3

(a) A Dense stimulus which contained 11 micro-patterns per row where  $\lambda_1 = 0.44$  degs (2.25c/d) and  $\sigma = 0.75 \lambda$ , and was presented with a short SOA of 60ms ( $\lambda_1$  is  $\lambda$  of the micro-patterns in the first exposure). Percentage errors in direction discrimination is shown as a function of the displacement (multiples of  $\lambda_1$ ). The continuous line shows results when the micro-patterns are unaltered between exposures of the motion sequence. The filled triangles show results when the micro-patterns in the second exposure had a spatial frequency of 1.125 c/d. The open circles show performance when the micro-patterns in the second exposure had a spatial frequency of 4.5c/d. (b) As (a) except the stimulus was sparsely populated with 5 micro-patterns per row, and presented with a long SOA of 140ms.

The "nonlinear" mechanism. The number of micro-patterns and the temporal offset (SOA) were selected for optimal performance of the "nonlinear" mechanism. That is, an SOA of 140ms, and a sparse pattern. The spatial frequency content of the micro-patterns was manipulated as above. The results are shown in figure 3b. The continuous line shows the results displayed in figure 2b i.e. no change in spatial frequency between the exposures of the motion sequence. The two dashed lines show results for when the spatial frequency of the second exposure differed by  $\pm$  one octave. In these cases performance is hardly affected, although there is a reduction in performance for small displacements. It is possible that in the "same spatial frequency" condition, there is still a small contribution from the spatial frequency tuned "quasi-linear" mechanism which is removed when it is no longer possible to use the spatial frequency content of the micro-pattern as a cue to correspondence. The "nonlinear" mechanism is shown to be able to detect correspondence between micro-patterns which differ in their internal structure. This mechanism can be characterized as an "envelope motion" detector, and as such

it should also be possible to change the orientation of the contents of the micro-patterns without inhibiting the detection of motion.

### 3.3 Orientation selectivity.

The "quasi-linear" mechanism. The number of micro-patterns and the temporal offset (SOA) were again selected for optimal performance of the "quasi-linear" mechanism. That is, an SOA of 60ms, and a dense pattern. In these experiments the spatial frequency was  $1c/d$ . In the first exposure of the motion sequence the cosine within the micro-patterns was orientated vertically (as in all other experiments). In the second exposure the micro-patterns were of the same size and frequency as in the first but the cosine component was orientated horizontally. The results are shown in figure 4a. The continuous line shows results similar to those shown in figure 2a, i.e. no change in orientation between the exposures of the motion sequence. The dashed line shows results for when the orientation of the second exposure differed by 90 degrees to that in the first. In this case performance is reduced to chance (50% errors) demonstrating that the "quasi-linear" mechanism is sharply tuned to orientation. Correspondence between spatial patterns differing by 90 degs in orientation is not achieved.

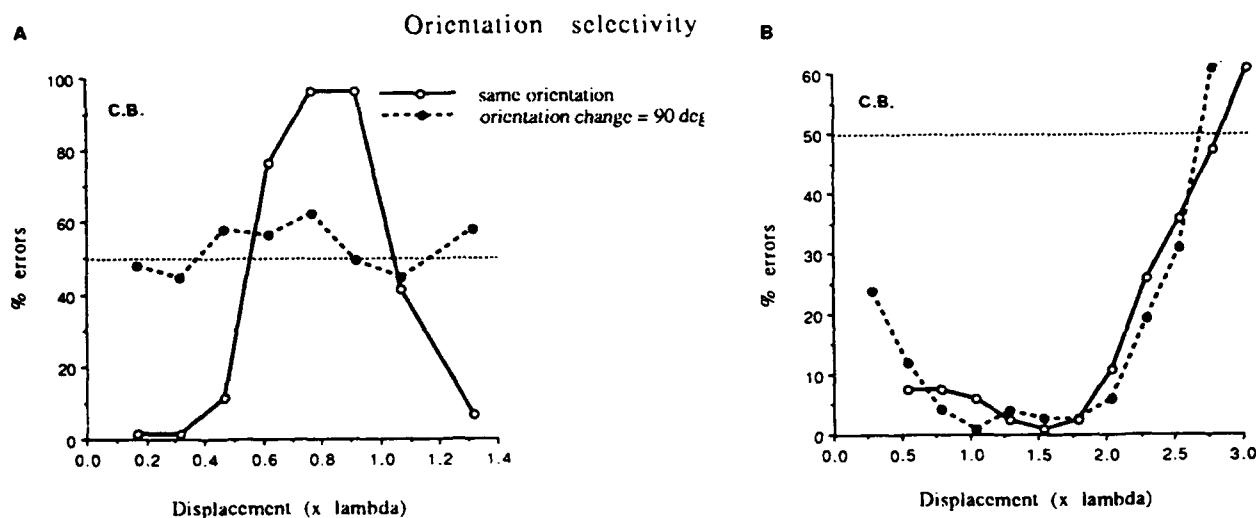


Figure 4.

(a) A dense stimulus was presented where  $\lambda = 1$  deg and  $\sigma = 0.75\lambda$  ( $\sigma_x = \sigma_y$ , so the envelope of the micro-patterns was circular), which gave micro-patterns 2.25 times larger than in the previous experiments. In this case a dense stimulus comprised of 5 micro-patterns per row. The stimulus was presented with an SOA of 60ms. The continuous line shows direction discrimination when the micro-patterns are unaltered between exposures. The dashed line shows performance when the orientation of the second exposure differed by 90 degs between exposures. (b) As (a) but a sparsely populated stimulus. For this size of micro-pattern ( $\lambda = 1$  deg) a sparse stimulus was comprised of 2 micro-patterns per row, SOA = 140ms.

The "nonlinear" mechanism. The number of micro-patterns and the temporal offset (SOA) were again selected for optimal performance of the "nonlinear" mechanism. That is, an SOA of 140ms, and a sparse pattern. The micro-patterns were manipulated as above. Performance for direction discrimination is shown as a function of the displacement in figure 4b. The continuous line shows the results similar to those shown in figure 2b i.e. no change in orientation between the exposures of the motion sequence. The dashed line shows results for when the orientation of the second exposure differed by 90 degrees. In this case performance is hardly affected. These results support the hypothesis that the "nonlinear" mechanism is able to make correspondence between micro-patterns which differ in their internal structure.

## 4. MODEL AND NEURAL SUBSTRATE.

We can characterize the two mechanisms in the following ways. The "quasi-linear" mechanism is fast (short SOAs) and narrowly tuned to spatial frequency and orientation. A suitable model would be an Elaborated Reichardt Detector (ERD)

as proposed by Van Santen and Sperling<sup>11</sup> (see introduction). The "nonlinear" mechanism can be characterized as an "envelope motion" detector. That is, it detects the motion of the contrast envelope of the stimulus (although there is no power at that scale in the Fourier domain) without making use of the internal structure of those envelopes. It is therefore, not tuned for orientation or spatial frequency. Furthermore it appears to have a longer latency (needs long SOAs), this also gains support from other studies<sup>25,26</sup>. This mechanism can be modeled by adding extra processing after the spatial frequency and orientation tuned filters, which characterize the input to the motion system<sup>18,19</sup>, and before the direction extracting unit. This extra processing has been suggested to be full wave rectification followed by smoothing<sup>4</sup>, which would account for our results.

The apparently longer latency of this mechanism could be accounted for by the extra processing involved in this mechanism. Alternatively this mechanism could be mediated by a different group of cells in the visual system which have slower temporal characteristics.

There is a wealth of anatomical and physiological evidence which suggests two distinct retinal-cortical processing streams; the parvocellular and the magnocellular pathways. The parvocellular pathway is postulated to mediate fine detail and colour processing whereas the magnocellular pathway is thought to mediate motion processing<sup>27,28,29,30,31,32</sup>. The magnocellular pathway has been shown to contain large fast conducting cells, whereas the parvo cellular pathway contains smaller slower cells (For a review see Kaplan, Lee and Shapley<sup>33</sup>). There are a number of reports which have investigated the ability of the slower parvocellular pathway to support motion perception by exploring the sensitivity of the colour system. They show that under limited conditions motion is detected, although the percept is often of poor quality<sup>34,35,36,37</sup>. We have shown that when colour vision detects motion it does so via the "nonlinear" mechanism described above<sup>38</sup>. This suggests that the parvocellular layer can support the "nonlinear" motion mechanism, and raises the possibility that in the luminance domain, this nonlinear mechanism could be mediated via the parvocellular pathway.

A further differentiating feature of parvo cells from magno cells is their respective contrast gain functions. Magno cells have been shown to have rapid contrast gain followed by saturation at relatively low contrast. Whereas the parvo cells have been shown to have slower contrast gain which does not saturate (if at all) until much higher contrast levels<sup>33,39</sup>. To gain insight as to whether the parvocellular pathway is a possible neural substrate for the "nonlinear" mechanism, we investigated the contrast gain functions for the "quasi-linear" and "nonlinear" mechanisms.

#### 4.1 Contrast gain functions.

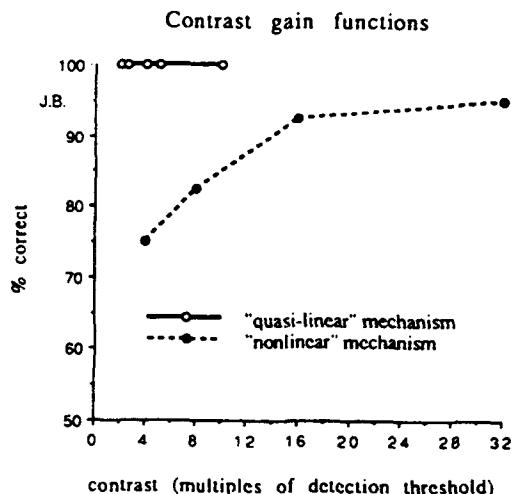
The number of micro-patterns and the temporal offset (SOA) were again selected for optimal performance of the "quasi-linear" mechanism. That is, an SOA of 60ms, and a dense pattern. The stimulus was presented in two frame apparent motion with a displacement of a quarter of a cycle of the wavelength of the cosine component of the Gabor micro-pattern, ( $0.25 \lambda$  is the optimal displacement for this stimulus, see figure 2a). To measure the contrast detection threshold, percentage correct for identifying the presence of the stimulus was measured with a method of constant stimuli for different contrast levels. Threshold was defined as 80% correct. The percentage errors in direction discrimination (for a displacement of  $0.25 \lambda$ ) was then measured for a range of higher contrasts. This was repeated for a stimulus with parameters optimized for the "nonlinear" mechanism. That is, an SOA of 140ms, a sparse stimulus, and the displacement was  $1.5 \lambda$ .

The percentage correct in direction discrimination are shown as a function of contrast (plotted in multiples of detection threshold) in figure 5. For the stimulus optimized for the "quasi-linear" mechanism, as soon as it was visible the direction of motion was detectable. However, for the "nonlinear" mechanism, performance was near chance when the contrast was near threshold, and showed improvement as the contrast was increased. Performance had not yet saturated at contrast levels of 30 times detection threshold. This indicates a lower contrast gain for the nonlinear mechanism than for the quasi-linear mechanism. Although the true contrast gain function for the "quasi-linear" mechanism was not obtained in these experiments (optimum performance was reached as soon as the stimulus was visible), it is clear that there is a significant difference between the two mechanisms in their dependence on the contrast of the stimulus. It has been documented elsewhere that the minimum displacement detectable, for a stimulus with parameters to which the quasi-linear mechanism would be sensitive, asymptotes at around  $10 \times$  threshold contrast<sup>40</sup>. Results which have been interpreted as supporting the hypothesis

that this mechanism is mediated via the magnocellular pathway. The low contrast gain shown for the "nonlinear" mechanism, which has not saturated at 30 times threshold, is characteristic of parvo cells. This evidence, together with the evidence from experiments in the colour domain<sup>38</sup> imply that (at least the initial stages) processing of the "nonlinear" mechanism is via the parvocellular pathway.

**Figure 5.**

The continuous line shows percentage correct for direction discrimination for a dense stimulus (11 micro-patterns per row) displayed with an SOA of 60ms as a function of the contrast of the stimulus. The displacement was optimal for this stimulus,  $0.25\lambda$  ( $\lambda = 0.44$  degs), see figure 2a. The dashed line shows performance for direction discrimination of a sparse stimulus (5 micro-patterns per row) displayed with an SOA of 140ms, as a function of the contrast of the stimulus. The displacement was  $1.5\lambda$  which was optimal for this stimulus, see figure 2b.



It is possible to postulate, that the two retina-cortical processing streams have a functional significance for motion processing, with the magnocellular pathway processing the "quasi-linear" mechanism, and the parvocellular pathway processing the "nonlinear" mechanism. The direction extraction for both mechanisms may be mediated by one of the known motion sensitive areas (eg. M.T.), but prior to that stage there is evidence for parallel processing of the two motion mechanisms.

## 5. CONCLUSIONS

Two mechanisms for the detection of motion have been demonstrated by the use of Gabor Kinematograms that maintain the same luminance correspondence under all conditions, but for which performance changes radically as temporal and spatial parameters are manipulated. These two mechanisms have been characterized as: a "quasi-linear" mechanism (which is fast, with a high contrast gain, and narrowly tuned for spatial frequency and orientation) and a "nonlinear" mechanism (that has a long latency, and is not tuned for orientation or spatial frequency<sup>41</sup>). It is postulated that the "quasi-linear" mechanism is mediated via the magnocellular retina-cortical pathway, and the "nonlinear" mechanism via the parvocellular retina-cortical pathway.

## 6. ACKNOWLEDGMENTS.

This research is supported by a grant from the Dutch Foundation for the Advancement of Pure Research (NWO) and a grant from the Canadian NSERC.

## 7. REFERENCES.

1. O.J. Braddick. "A short range process in apparent motion." *Vision Research*, 14, 519-527. 1974.
2. O.J. Braddick. "Low-level and high-level processing in apparent motion." *Philosophical Transactions of the Royal Society of London B*, 290, 137-151. 1980.

3. S. Anstis. "The perception of apparent motion." *Philosophical Transactions of the Royal Society of London, B290*, 153-168. 1980.
4. C. Chubb, & G. Sperling. "Drift-balanced random stimuli: a general basis for studying non-fourier motion perception." *Journal of the Optical Society of America A*, 5, 1986-2006. 1988.
5. J.T. Petersik, & A. Pantle. "Factors controlling the competing sensations produced by a bistable stroboscopic motion display." *Vision Research*, 19, 143-154. 1979.
6. C.L. Baker, & O.J. Braddick. "The basis of area and dot number effects in random dot motion perception." *Vision Research*, 22, 1253-1259. 1982.
7. C.L. Baker, & O.J. Braddick. "Temporal properties of the short-range process in apparent motion." *Perception*, 14, 181-192. 1985.
8. C.L. Baker, & O.J. Braddick. "Eccentricity dependent scaling of the limits for short range apparent motion perception." *Vision Research*, 25, 803-812. 1985.
9. M. Wertheimer. "Untersuchungen zur Lehre der Gestalt." *Psychologische Forschung*, 4, 301-350. 1912.
10. J.P.H. van Santen, & G. Sperling. "A temporal covariance model of motion perception." *Journal of the Optical Society of America A*, 1, 451-473. 1984.
11. J.P.H. van Santen, & G. Sperling. "Elaborated Reichardt detectors." *Journal of the Optical Society of America A*, 2, 300-321. 1985.
12. E.H. Adelson, & J.R. Bergen. "Spatiotemporal energy models for the perception of motion." *Journal of the Optical Society of America A*, 2, 284-299. 1985.
13. A.B. Watson, & A.J. Ahumada Jr. "Model of human visual motion sensing." *Journal of the Optical Society of America A*, 1, 322-342. 1985.
14. P. Cavanagh, & G. Mather. "Motion: the long and the short of it." *Spatial Vision*, 4, 103-129. 1989.
15. P. Cavanagh. "Short-range vs long-range motion: Not a valid distinction." *Spatial Vision*, 5, 303-309. 1991.
16. A. Pantle, S. Lehmkuhle, & M. Caudill. "On the capacity of directionally selective mechanisms to encode different dimensions of moving stimuli." *Perception*, 7, 261-267. 1978.
17. K. Turano, & A. Pantle. "Discontinuity limits for the generation of visual motion aftereffects with sine- and square-wave gratings." *Journal of the Optical Society of America A*, 2, 260-266. 1985.
18. R.L. DeValois, D.G. Albrecht, & L.G. Thorell. "Spatial frequency selectivity of cells in macaque visual cortex." *Vision Research*, 22, 545-559. 1982.
19. C.L. Baker, & M.S. Cynader. "Spatial receptive field properties of direction selective neurons in cat striate cortex." *Journal of Neurophysiology*, 55, 1136-1152. 1986.
20. J.C. Boulton, & R.F. Hess. "The optimal displacement for the detection of motion." *Vision Research*, 30, 1101-1106. 1990.

21 E.L. Cameron, C.L. Baker & J.C. Boulton. "Spatial frequency selectivity mechanisms underlying the motion aftereffect." *Vision Research*, 32, 561-568. 1992.

22. J.C. Boulton & C.L. Baker. "Motion detection is dependent on spatial frequency not size." *Vision Research*, 31, 77-87. 1991.

23. J.C. Boulton & C.L. Baker. "Different parameters control motion perception above and below a critical density." *Vision Research*, in press. 1993.

24. J.C. Boulton & C.L. Baker. "Dependence on stimulus onset asynchrony in apparent motion: evidence for two mechanisms." *Vision Research*, in press. 1993.

25. M.A. Georgeson & M.G. Harris. "The temporal range of motion sensing and motion perception." *Vision Research*, 30, 615-619. 1990.

26. A.M. Derrington, D.R. Badcock & G.B. Henning. "Discriminating the direction of second-order motion at short stimulus durations." *Vision Research*, in press. 1993.

27. S.M. Zeki. "Functional specialization in the visual cortex of the rhesus monkey." *Nature*, 274, 423-428. 1978.

28. L.G. Underleider & M. Mishkin. "Two cortical visual streams." In D.J. Ingle, M.A. Goodale, & R.J.W. Mansfield (Eds), *Analysis of visual behaviour* (pp 549-586). Cambridge, Mass.: MIT Press. 1982.

29. D.C. Van Essen & J.H.R. Maunsell. "Hierarchical organization and functional streams in the visual cortex." *Trends in Neuroscience*, 6, 370-375. 1983.

30. D.H. Hubel & M.S. Livingstone. "Segregation of form color and stereopsis in primate area 18." *Journal of Neuroscience*, 7, 3378-3415. 1987.

31. E.A. De Yoe & D.C. Van Essen. "Concurrent processing streams in monkey visual cortex." *Trends in Neuroscience*, 5, 219-226. 1988.

32. S.M. Zeki, J.D.G. Watson, C.J. Lueck, K.J. Fritson, C. Kennard & R.S.J. Frackowiak. "A direct demonstration of functional specialization in human visual cortex." *Journal of Neuroscience*, 11, 641-649. 1991.

33. E. Kaplan, B.B. Lee & R.M. Shapley. "New views of primate retinal function." In *Progress in Retinal Research*, 9, 273-335. Pergamon Press (Oxford). 1989.

34. P. Cavanagh, C.W. Tyler & O. Favreau. "Perceived velocity of moving chromatic gratings." *Journal of the Optical Society of America*, 1, 893-899. 1984.

35. D.T. Lindsey, & D. Teller. "Motion at isoluminance: discrimination/detection ratios for moving isoluminant gratings." *Vision Research*, 30, 1751-1761. 1990.

36. K.T. Mullen & J.C. Boulton. "Absence of smooth motion perception in color vision." *Vision Research*, 32, 483-488. 1992.

37. K.R. Dobkins & T.D. Albright. "What happens if it changes color when it moves?: psychophysical experiments on the nature of chromatic input to motion detectors." *Vision Research*, 33, 1019-1036. 1993.

38. J.C. Boulton, C.L. Baker, & K.T. Mullen. "Nonlinear motion perception at isoluminance." *Investigative Ophthalmology and Visual Science*, 34, supplemental. 1033. 1993.

39. A.M. Derrington & P.Lennie. "Spatial and temporal contrast sensitivities of neurons in lateral geniculate nucleus of macaque." *Journal of Physiology*, 357, 219-240. 1984.

40. K. Nakayama & G.H. Silverman. "Detection and discrimination of sinusoidal grating displacements." *Journal of the Optical Society of America, A 2*, 457-473. 1985.

41. P. Werkhoven, G. Sperling, & C. Chubb. "The dimensionality of texture-defined motion: a single channel theory." *Vision Research*, 33, 463-485. 1993.

**Motion mechanisms based on non-linear spatial filters have lower temporal resolution.**

Andrew M Derrington

Department of Physiological Sciences; The Medical School, University of Newcastle upon Tyne,  
Newcastle upon Tyne, NE2 4HH, UK

**ABSTRACT**

Analysis of the motion of spatial patterns may be accomplished by analysing the spatio-temporal variations caused when a spatially varying luminance waveform moves over the detector surface. Non-linear transformations (such as squaring) of the input signal may give rise to a signal (a "distortion product") that varies on a different spatial scale from that of the original, and can thus give rise to a motion signal that is processed by a different set of spatiotemporal filters.

Experiments with patterns made by adding together two sinusoidal gratings, differing in spatial frequency or orientation and in temporal frequency, show that the human visual system can analyse the motion of the "difference-frequency" distortion products that would be introduced by squaring, and thus must contain mechanisms that use some non-linear transformation of this sort. This raises a question: is the non-linearity simply an inherent part of the transduction process, or do separate linear and non-linear motion analysers exist? We find that performance in motion discrimination tasks that require non-linear analysers declines rapidly for stimulus durations less than about 200 msec, and for temporal frequencies greater than about 1 Hz, whereas discriminations based on linear analyses are reliable and correct at durations down to 20 msec, and at temporal frequencies over 10 Hz.

This suggests that the linear and non-linear motion analysers are different.

**INTRODUCTION**

**Low-level and high-level motion analysis: does squaring require a high-level mechanism?**

Information about motion can be extracted from a visual image by a variety of different processing strategies. These different strategies may or may not reflect modes of operation of different sets of visual mechanisms. For example, it is now general practice to distinguish between low-level mechanisms, which calculate a motion signal by spatio-temporal correlation (or Fourier analysis) of the raw luminance values in the image<sup>1-5</sup>, and high-level mechanisms, which require the image to be processed in some way to extract data which can then be used for such an analysis<sup>6-10</sup>. The low-level mechanisms work by filtering motion-energy present in the image<sup>1</sup>; the high level mechanisms can be thought of as filtering motion energy introduced into the image by a preprocessing stage<sup>10</sup>. This paper considers whether a very simple form of pre-processing (squaring), might actually be inherent in the mechanisms for the low-level analysis of motion. The reasons for considering this possibility originate in psychophysical experiments, but draw further support from physiological studies.

First, it has long been recognised that non-linear transduction could add a component proportional to the square of the image to its internal representation. This would introduce a low spatial frequency *distortion product* into the internal representation of patterns. The signal producing the distortion product consist of a large-scale (low spatial frequency) spatial modulation of the contrast of a higher frequency carrier<sup>11, 12</sup>. Naturally movement of the modulating signal would, through motion of the low spatial frequency distortion product, give rise to a motion energy signal in the non-linearly transformed internal representation of the image<sup>13</sup>, although it would not generate any net motion energy in the image itself.

Physiological experiments show that, in the cat at least, an appropriate distortion product arises at a point in the visual pathway before direction-selective spatial filtering appears. Recordings from X-cells in the LGN of the cat show that their responses to moving amplitude-modulated patterns contain a component proportional to the square of the local contrast, which in turn would generate a signal proportional to the motion of the modulating waveform at the input to the striate cortex<sup>14, 15</sup>. Thus it seems that the direction-selective spatio-temporal filters early in the visual pathway have available at

their inputs distortion products which would enable them to signal the direction of motion of spatial variations in contrast. In other words, motion of the contrast "envelope" might in principle be extracted by a low-level mechanism. In this paper I describe experiments which address this question by studying the effect of varying the duration and the temporal frequency of complex moving patterns.

#### **Decreasing duration favours "low-level" motion-analysis mechanisms.**

It seems fairly well established that reducing the stimulus duration, or increasing its temporal frequency favours low-level motion-analysis mechanism over high-level mechanisms. The effect has been widely studied in a variety of ambiguous stimuli in which low-level and high-level systems signal different motions. Anstis<sup>16</sup> showed that the motion of a grating that alternated periodically between two different orientations could be perceived in two ways. With long intervals between the alternations, the motion was seen as global rotation of the grating, a "high-level" percept because in order to perceive global rotation, one must first extract the form of the grating. With shorter intervals between the position changes, the motion was seen as a local change at the spatial intersections within the two gratings. This percept was presumably mediated by a low-level motion system, since it would be produced by a spatio-temporal correlation of the local illuminance values of the image. Similar changes from "high-level" to "low-level" motion percepts have been reported as a consequence of reducing the interstimulus interval in apparent motion displays that could be interpreted either as motion of individual local elements of the pattern (low-level) or as a coordinated movement of the whole pattern (high-level)<sup>8, 17</sup>.

It is now generally assumed that one can isolate the low-level motion system simply by using brief presentations. For example Yo and Wilson<sup>18</sup> show that the way in which the perceived axis of motion of a complex 2-dimensional moving pattern changes with its duration can be explained by a model in which the signals from low-level and high-level motion analyses are combined<sup>19</sup>. The perceived axis of motion changes with duration because the high-level analysis is assumed to take longer than the first-order analysis<sup>18</sup>. Consequently, in briefly presented stimuli, only the low-level mechanisms are able to provide signals, but as the stimulus duration is increased, signals from the slower high-level mechanisms gradually become available and influence the motion percept. However, in Wilson et al.'s<sup>19</sup> model it is assumed that the squaring process which extracts the spatial contrast envelope takes place after the first stage of direction-selective filtering.

The fact that a squaring process of this sort occurs early in the visual pathway<sup>14, 15</sup> makes it important to compare the temporal properties of the psychophysical mechanisms which extract the motion of luminance patterns with those of the mechanisms that extract the motion of contrast patterns. To that end we have chosen to investigate the effects of exposure-duration and temporal frequency on the detectability of the motion of a very simple stimulus which contains a moving contrast envelope. We have used a "beat" pattern, which is formed by adding together two gratings of about 6 c/deg which differ in frequency by about 1 c/deg. The pattern appears as a spatially periodic variation in the contrast of a grating whose frequency is the mean of the frequencies of the two components. The period of the variation in contrast is equal to the difference between the frequencies of the two components; and, if they are made to move in opposite directions with equal temporal frequencies, the low-frequency contrast variation moves but the high-frequency "carrier" grating remains stationary<sup>20</sup>. For comparison we also study sensitivity to motion of simple luminance patterns of low spatial frequency (sinusoidal gratings) and compound patterns made by summing gratings of different orientations.

## **METHODS**

### **Stimuli**

Patterns were generated using a RGB framestore that was part of a purpose built display controller, the Cambridge Research Systems VSG 2/1<sup>21</sup>, and displayed on a Joyce Electronics monitor with a P4 (bluish white) phosphor. The 3 DAC outputs of the framestore were summed with different gains to give more precise control of contrast<sup>22</sup>. On each frame of the display (frame frequency 180 Hz or 120 Hz) a moving sinusoidal grating was presented within a circular patch, the diameter of which subtended 5° at the 2-m viewing distance. The mean luminance of the display was 47 cd.m<sup>-2</sup>; the illuminated area subtended 7.7° horizontally by 6.4° vertically, and it had a dark surround. The room was dimly illuminated.

Normally two different patterns were interleaved, each member of the pair being presented on alternate frames. Two different stimulus pairings were used: 1) a vertical grating paired with a blank field, and two spatially-superimposed

gratings of different spatial frequencies, to produce a "beat" pattern, illustrated in Fig 2b, or of different orientations, to produce a "plaid" pattern<sup>23</sup>. The grating pattern is described by the equation

$$L(x,t) = L_m \{ 1 + C \cos [2\pi(fx + gt) + \phi] \} \quad (1)$$

where  $L_m$  is the mean luminance,  $C$  is the contrast,  $f$  is the spatial frequency,  $g$  is the temporal frequency and  $\phi$  is a phase term. The beat or plaid pattern was made by adding together two sinusoidal gratings of different spatial frequencies or orientations, and is described by the general equation

$$L(x,y,t) = L_m \{ 1 + C \cos [2\pi(u_1x + v_1y + gt) + \phi_1] + C \cos [2\pi(u_2x - v_2y + gt) + \phi_2] \}. \quad (2)$$

In the case of the beat pattern,  $v_1 = v_2 = 0$ ,  $u_1 = (f_c + f_e)$ ,  $u_2 = (f_c - f_e)$  and  $g_1 = -g_2 = g$ . In the case of the plaid pattern,  $u_1 = u_2 = u$ ,  $v_1 = -v_2 = v$  and  $g_1 = g_2 = g$ , so the two components were oriented symmetrically  $\pm \arctan(v/u)$  from the vertical, and their spatial frequency was  $\sqrt{(u^2 + v^2)}$ .

In the case of the beat pattern equation (2) can be rewritten as

$$L(x,t) = L_m \{ 1 + 2C \cos(2\pi f_c x + \pi gt + \phi_c) \cos(2\pi f_c + \phi_c) \} \quad (3)$$

expressing the pattern as the product of a moving sinusoidal envelope, of spatial frequency  $f_e$  and temporal frequency  $g/2$ , and a static sinusoidal carrier of spatial frequency  $f_c$ . However the spatial modulation in the contrast of the carrier has a periodicity twice that of its envelope because we are unable to distinguish the positive and negative peaks of the modulating waveform, and so we refer to the spatial frequency of the beat as  $f_b$ , where  $f_b = 2f_e$ , and to its temporal frequency as  $g$ .

In one experiment we used an amplitude-modulated grating with a moving modulation envelope and a static carrier described by the equation

$$L(x,t) = L_m \{ 1 + C \{ 1 + m \sin(2\pi f_m x + 2\pi gt + \phi_m) \} \sin(2\pi f_c + \phi_c) \}. \quad (4)$$

The carrier contrast,  $C$ , was 0.1; the carrier frequency,  $f_c$ , was 5 c/deg; the modulation spatial frequency,  $f_m$ , was 1 c/deg; the modulation temporal frequency was the independent variable of the experiment and the modulation depth was varied to measure threshold.

All patterns were modulations of luminance without changes in chromaticity, and were presented with abrupt onset and offset. The spatial frequency of the grating used was 0.93 cycles/degree (c/deg), that of the components of the plaid pattern 6.0 c/deg; they were orientated  $\pm 81^\circ$  from vertical. The spatial frequencies of the components of the beat pattern were 5.4 c/deg and 6.3 c/deg. This gave a spatial frequency of 0.93 c/deg for the beat pattern. The plaid pattern had a horizontal period of 1.1 degrees, but an apparent horizontal period of 0.54 degrees, half that of the beat pattern. In preliminary experiments, we found that raising the beat frequency or increasing the angle between the components of the plaid pattern caused a tendency for the moving patterns to break up.

The sinusoidal grating patterns were generated by storing lookup-table index values in the parts of display memory that were displayed as the circular patch on alternate frames. The memory locations corresponding to the rest of the visible screen area contained the index of the lookup table entry containing the mean luminance. Separate lookup tables, each containing 251 gamma-corrected luminance values corresponding to a full cycle of a sine-wave of contrast 0.1, were maintained for each pattern. Thus the part of display memory representing each pixel contained a number which indicated the phase of the sinusoid at that point in the picture. The lookup table was used to convert that phase into the three numbers which, when loaded in the 8-bit DACs, gave the luminance required at that phase for a sinusoidal grating of contrast 0.1. Because each grating was interleaved either with another, or with a blank field, the time-average contrast of each grating was always 0.05.

Each pattern could be made to move smoothly within its circular patch by loading a new lookup table each time the pattern was displayed (90 times per second). The smallest unit of phase shift was 1/251 cycles, giving a temporal frequency

resolution of 0.36Hz independent of the spatial frequency of the pattern. To obtain finer resolution of temporal frequency at the expense of smoothness of movement, phase shifts were made in units of 1/2008 cycles (1/8 waveform samples) and rounded down to the nearest whole step on each frame. Thus the slowest possible motion, 0.045 Hz, would result in a phase shift once every 8 frames. The lack of smoothness in the motion was not noticeable.

### Subjects

Two or three practised observers, at least one of whom knew nothing of the theoretical background, provided data for each experiment. They viewed the screen without head restraint, and with natural pupils and accommodation. They were given a fixation mark, and were instructed to fixate. They wore their prescribed spectacle corrections. In all experiments results from different observers were similar.

### Procedures

A temporal two-alternative forced-choice (2-AFC) paradigm was used in conjunction with the method of constant stimuli to obtain psychometric functions (50 observations per point) relating performance in direction-discrimination tasks to stimulus duration. Each trial was initiated by a key-press, and consisted of two temporal intervals signalled to the observer by bursts of audible noise. During one interval, chosen at random, a pattern was presented moving to the left, during the other interval the same pattern was presented moving to the right. The Observer's task was to signal, by pressing a key-switch, the interval in which the pattern had moved to the left. Observers were given no feedback as to the correctness of their responses on individual trials.

The stimulus to be presented on each trial was selected at random from the set of six used for the current block of trials; with the constraint that no stimulus was presented for the  $n^{\text{th}}$  time until all stimuli had been presented  $n-1$  times. A computer (Tandon PCA 20), containing the VSG2/1, controlled the selection, generation and display of stimuli, and the recording of responses.

## RESULTS AND DISCUSSION

### Discrimination of direction of motion of stimuli of different durations

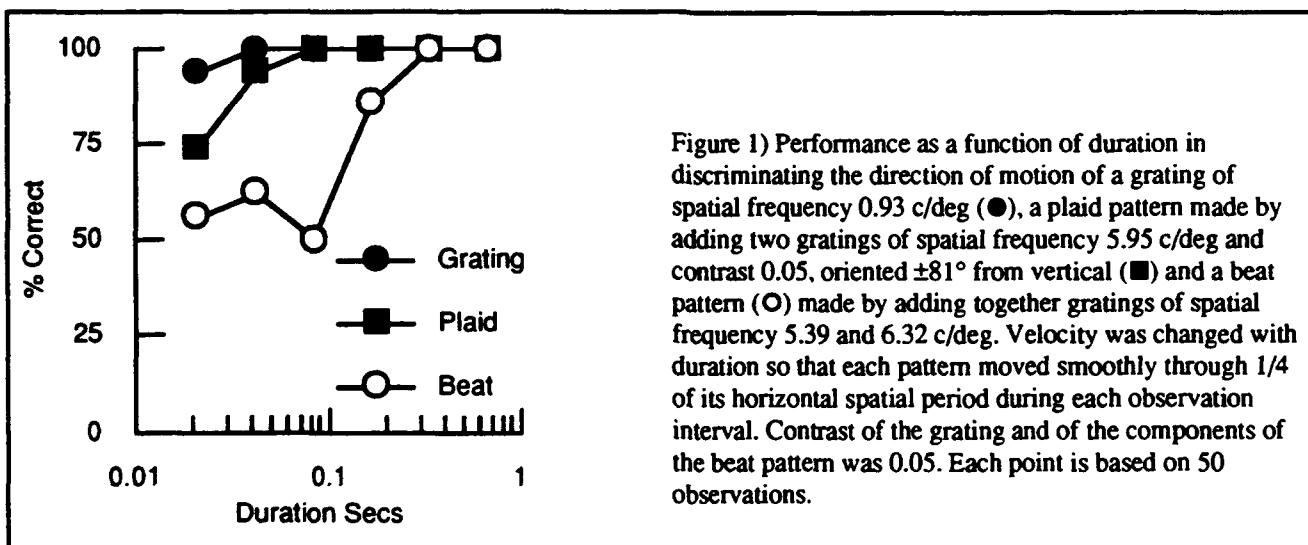


Figure 1 shows the performance of an observer discriminating the direction of motion of the three different types of pattern plotted as a function of the stimulus duration. The speed of the moving pattern covaried with the duration, so that the pattern always moved through 1/4 of its spatial period during each presentation. There is a marked difference between the results obtained with the different patterns. With the sinusoidal grating performance is essentially perfect at all durations

from 0.022 seconds. Performance with the plaid is not quite as good, but even at the shortest duration, it is approximately 75% correct, whereas the motion of the beat cannot be discriminated at short durations; and reaches 75% correct at a duration of about 200 msec. In order for the observers to see the motion of the beat it must be presented for about ten times as long as is required to see motion of the other two patterns. This result suggests that the mechanism which signals the motion of the beat is in some sense more sluggish than the mechanism which signals the motion of the grating.

#### Performance as a function of temporal frequency.

In the experiments described above the stimulus speed was adjusted so that the pattern moved through 1/4 of its apparent spatial period during the observation interval, with the consequence that its speed (or temporal frequency) was inversely related to its duration. At the shortest duration, the nominal temporal frequency was approximately 12 Hz (although the brief duration will have ensured that the stimulus really contained a broad band of temporal frequencies centred on this value). Thus there is a possibility that the decline in direction-of-motion discrimination performance at short durations could be a consequence of the increase in temporal frequency rather than the reduction in duration. To test this we studied performance as a function of stimulus temporal frequency at a number of durations using the same three stimuli. The results are shown in Figure 2.

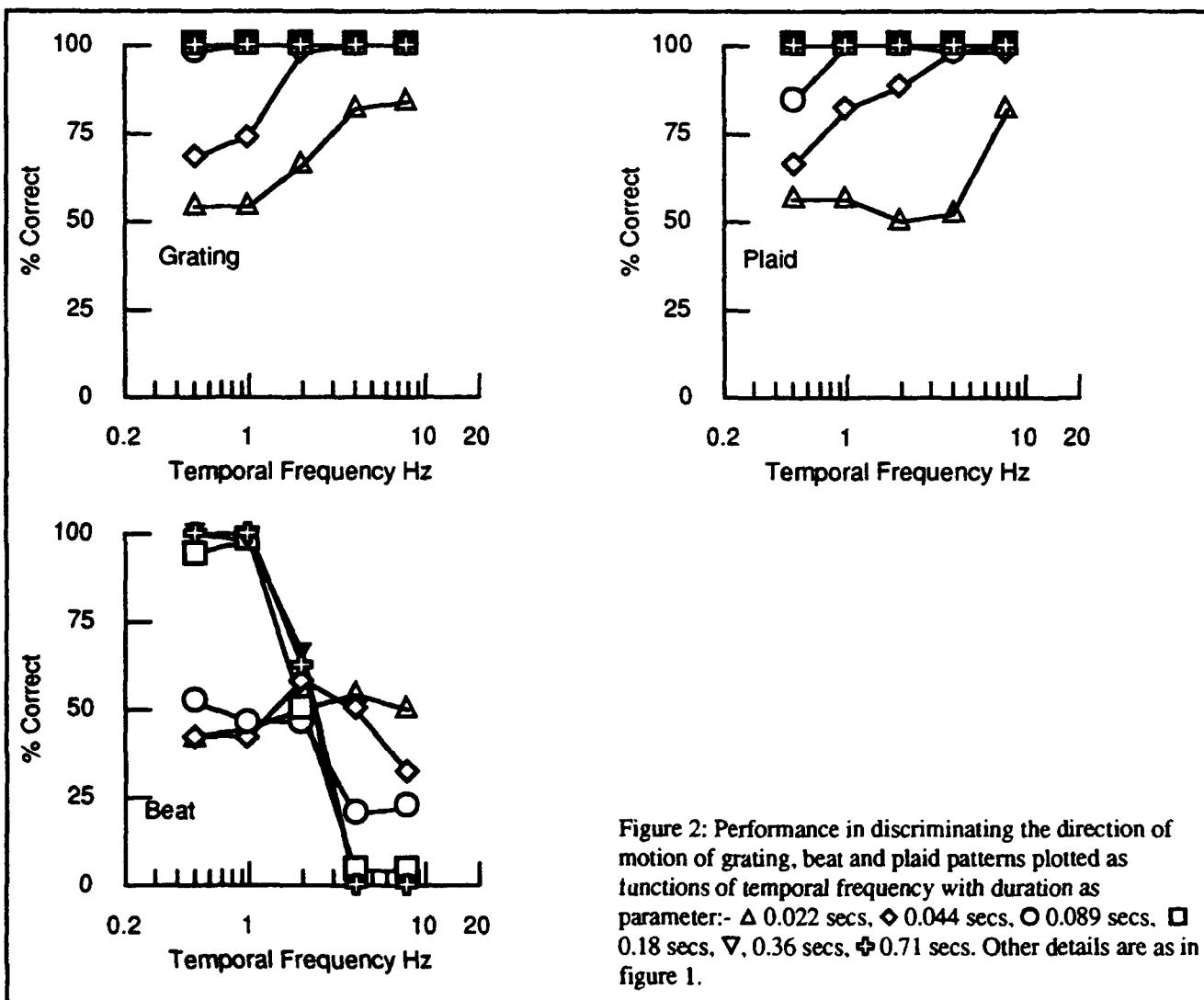


Figure 2 shows performance in discriminating direction of motion of the different stimuli as a function of the stimulus temporal frequency, with duration as a parameter. With the sinusoidal grating performance is best overall, and improves with both duration and temporal frequency within the range of temporal frequencies studied (0.5-8 Hz). Perfect

performance is obtained for all speeds at all durations above 40 msec. The same gradual improvement with both speed and duration is obtained with the plaid pattern. Except for the briefest stimulus duration, performance is above threshold at all except the slowest speeds.

The results obtained using the beat pattern are completely different. First, performance remains at or below chance for the stimulus durations below 0.1 seconds. Then, at longer durations, it becomes a monotonically declining function of temporal frequency. Perfect performance is obtained at frequencies up to 1 Hz, but the rapid decline in brings performance down to chance levels at 2 Hz, and at higher frequencies performance is close to zero. At temporal frequencies above 2 Hz observers see reversed motion.

The reversal of the motion percept at high temporal frequencies occurs in all observers. It is almost certainly due to the fact that the beat stimulus gives rise to a slightly stronger motion signal in the linear mechanisms selective for the direction opposite that of the beat's motion. It contains two gratings of slightly different spatial frequency moving in opposite directions at the same temporal frequency. Although this arrangement almost completely cancels the net motion energy in the stimulus, the visual system is slightly more sensitive to the component with the lower spatial frequency particularly at high temporal frequencies<sup>24, 25</sup>. The beat moves in the direction opposite that in which the lower spatial frequency component moves<sup>20</sup> so it seems likely that reversed motion percept is caused by the observers' simply responding to the motion of the more visible component of the beat. In fact the percept of reversed motion at high temporal frequencies seems to begin at shorter durations than does the percept of forward motion seen at low temporal frequencies: at a duration of 0.89 seconds performance at the two highest temporal frequencies is below 25% correct.

#### **Detection and discrimination of direction of motion of the envelope of an amplitude modulated grating.**

Two points arise from the results shown figure 2. First, these results may underestimate the high temporal frequency performance of the mechanism which signals motion of contrast variations, because, as the temporal frequency rises, the imbalance in sensitivity to the two component gratings will become more important, and the reversed motion signal generated by linear mechanisms will grow. Second, the results give no clue about whether the poor performance results from an inability to detect the spatial envelope or from an inability to discriminate its direction of motion. To resolve both of these questions we measured modulation thresholds for detecting and for discriminating the direction of motion of the envelope of an amplitude modulated grating. In this stimulus an imbalance still occurs, but it is much less noticeable because the stationary carrier has much higher contrast. The results (reciprocals of threshold contrast modulations) are shown in figure 3, with contrast sensitivity measurements made using a sinusoidal grating of the same spatial frequency as the modulation envelope shown for comparison.

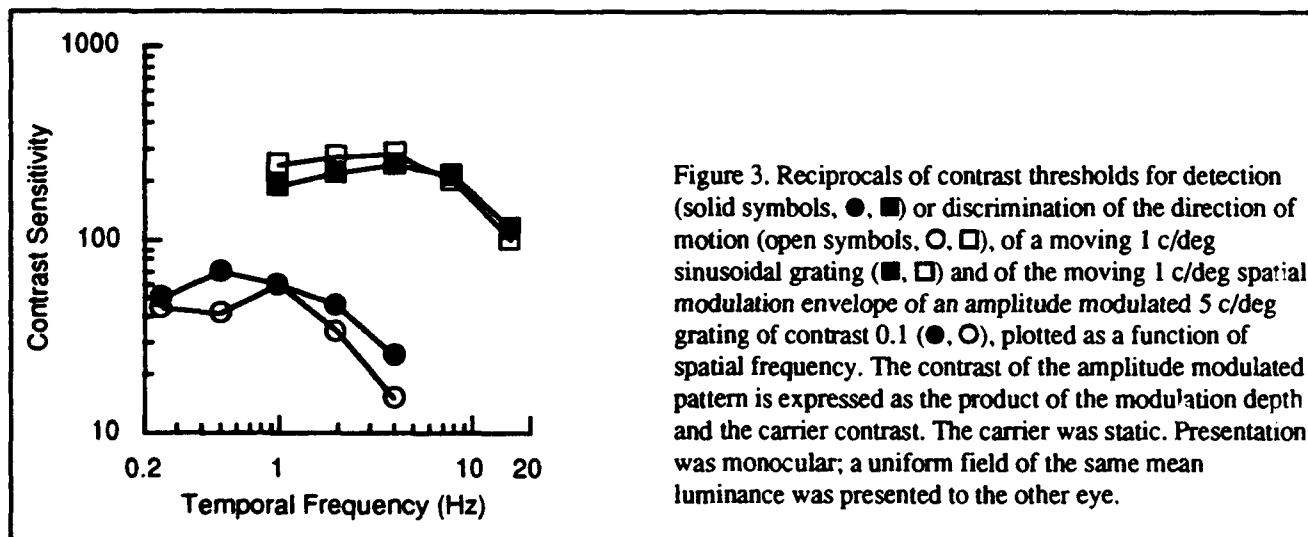


Figure 3. Reciprocals of contrast thresholds for detection (solid symbols, ●, ■) or discrimination of the direction of motion (open symbols, ○, □), of a moving 1 c/deg sinusoidal grating (■, □) and of the moving 1 c/deg spatial modulation envelope of an amplitude modulated 5 c/deg grating of contrast 0.1 (●, ○), plotted as a function of spatial frequency. The contrast of the amplitude modulated pattern is expressed as the product of the modulation depth and the carrier contrast. The carrier was static. Presentation was monocular; a uniform field of the same mean luminance was presented to the other eye.

The form of the relationship between sensitivity and temporal frequency is quite different for the two stimuli, but for each stimulus the performance in the two tasks is fairly similar. In both the detection and the direction discrimination task sensitivity to the amplitude-modulation envelope begins to fall off at very low temporal frequencies, 0.5 and 1 Hz.

respectively. At 4 Hz sensitivity is substantially less than half the peak sensitivity. Sensitivity in the direction discrimination task appears to fall significantly faster than it does in the detection task. Further, the mechanism detecting the sinusoidal grating clearly has substantially higher temporal resolution: peak sensitivity occurs at 4 Hz for both tasks, and at 16 Hz the sensitivity has fallen only by about half its best value. These differences could not simply result from differences in gain between the mechanism detecting the envelope and that detecting the grating. They indicate a substantial difference in temporal frequency tuning between the two mechanisms, suggesting that they are of radically different types.

## GENERAL DISCUSSION AND CONCLUSIONS

The results leave no doubt that the non-linear mechanisms which analyse the motion of spatial variations in contrast have much worse temporal resolution than do linear motion mechanisms. This suggests that if indeed the non-linear processing necessary to introduce motion energy into such patterns takes place at a low level in the visual system, as would be consistent with the physiological data showing quadratic distortion products in geniculate X-cells<sup>15</sup>, then the process that generates the distortion product must have much worse temporal resolution than do linear receptive field mechanisms. An alternative possibility, consistent with the psychophysical data, would be that the analysis of the motion of spatial variations in contrast takes place in a different kind of mechanism with lower temporal resolution than linear motion analysis mechanisms.

## ACKNOWLEDGEMENTS

This research was supported by SERC grants GR/G 00730 and GR/G 07982. I thank D.R. Badcock, M.J. Cox and E.C. Colombo for collecting the data, and G.B. Henning for helpful discussions.

## REFERENCES

1. Adelson, E.H., and J.R. Bergen, "Spatiotemporal energy models for the perception of motion," *Journal of the Optical Society of America* 2.2 (1985): 284-299.
2. Braddick, O., "A short-range process in apparent motion," *Vision Research* 14 (1974): 519-527.
3. Watson, A.B., and A.J. Jr. Ahumada, "Model of human visual-motion sensing," *Journal of the Optical Society of America* A2.2 (1985): 322-342.
4. van Santen, J.P.H., and G. Sperling, "Temporal covariance model of human motion perception," *Journal of the Optical Society of America* A1.5 (1984): 451- 473.
5. Reichardt, W., "Autocorrelation, a principle for the evaluation of sensory information by the central nervous system..," *Sensory Communication*, ed. W.A. Rosenblith. (New York: Wiley, 1961) 303-317.
6. Ullman, S., *The interpretation of visual motion.*, *The MIT Press series in Artificial Intelligence*, ed. (Cambridge, Massachusetts and London, England: MIT Press, 1979) 229.
7. Anstis, S.M., "The perception of apparent movement..," *Philosophical Transactions of the Royal Society of London* B290 (1980): 153-168.
8. Braddick, O.J., "Low-level and high-level processes in apparent motion," *Philosophical Transactions of the Royal Society of London* 290 (1980): 137-151.
9. Chubb, C., and G. Sperling, "Two motion perception mechanisms revealed through distance driven reversal of apparent motion," *Proceedings of the National Academy of Sciences* 86 (1989): 2985-2989.
10. Cavanagh, P., and G. Mather, "Motion: The long and short of it," *Spatial Vision* 4.2/3 (1989): 103-129.
11. Henning, G.B., B.G. Hertz, and D.E. Broadbent, "Some experiments bearing on the hypothesis that the visual system analyses spatial patterns in independent bands of spatial frequency," *Vision Research* 15 (1975): 887-897.
12. Burton, G.J., "Evidence for non-linear response processes in the human visual system from measurements on the thresholds of spatial beats" *Vision Research* 13 (1973): 1211-1225.
13. Badcock, D.R., and A.M. Derrington, "Detecting the displacement of periodic patterns," *Vision Research* 25 (1985): 1253-1258.
14. Derrington, A.M., "Amplitude-modulated gratings mask the responses of X-cells in cat LGN to gratings of the modulation frequency," *Journal of Physiology* 391 (1987): 39P.
15. Derrington, A.M., "Distortion products in geniculate X cells: a physiological basis for masking by spatially modulated gratings?," *Vision Research* 27 (1987): 1377-1386.

16. Anstis, S.M., "Phi movement as a subtraction process," *Vision Research* 10 (1970): 1411-1430.
17. Petersik, J.T., and A. Pantle, "Factors controlling the competing sensations produced by a bistable stroboscopic motion display," *Vision Research* 19 (1979): 143-154.
18. Yo, C., and H.R. Wilson, "Perceived direction of moving two-dimensional patterns depends on duration, contrast and eccentricity," *Vision Research* 32.1 (1992): 135-147.
19. Wilson, H.R., V.P. Ferrera, and C. Yo, "A psychophysically motivated model for two-dimensional motion perception," *Visual Neuroscience* 9 (1992): 79-97.
20. Derrington, A.M., and D.R. Badcock, "Separate detectors for simple and complex grating patterns?," *Vision Research* 25.12 (1985): 1869-1878.
21. Derrington, A.M., and M. Suero, "Motion of complex patterns is computed from the perceived motions of their components," *Vision Research* 31.1 (1991): 139-149.
22. Pelli, D.G., and L Zhang, "Accurate control of contrast on microcomputer displays," *Vision Research* 31.7/8 (1991): 1337-1350.
23. Adelson, E.H., and J.A. Movshon, "Phenomenal coherence of moving visual patterns," *Nature* 300.5892 (1982): 523-525.
24. Robson, J.G., "Spatial and temporal contrast sensitivity functions of the visual system," *Journal of the Optical Society of America* 56 (1966): 1141-1142.
25. Henning, G.B., "Spatial-frequency tuning as a function of temporal frequency and stimulus motion," *Journal of the Optical Society of America* A5 (1988): 1362- 1373.

# Fidelity Metrics and the Test-Pedestal Approach to Spatial Vision

Stanley A. Klein

School of Optometry, UC Berkeley  
Berkeley, California 94720

## ABSTRACT

This paper has three parts. Part 1, the Introduction, contains musings on the title of this conference, "Computational Vision Based on Neurobiology" held at Asilomar. One of the musings is that progress has been slow in computational vision because very difficult problems are being tackled before the simpler problems have been solved. Part 2 is about one of these simpler problems in computational vision that is largely neglected by computational vision researchers: the development of a fidelity metric. This is an enterprise perfectly suited for computational vision with the side benefit of having spectacular practical implications. Part 3 discusses the research my colleagues and I have been pursuing for the past several years on the Test-Pedestal approach to spatial vision. This approach can be helpful as a guide for the development of a fidelity metric. A number of experiments using this approach are discussed. These examples demonstrate both the power and the pitfalls of the Test-Pedestal approach.

## 1. INTRODUCTION.

The conference on which this book is based was centered on how neurobiology could aid computational vision. There were many excellent presentations demonstrating the progress being made in neurobiology. I will always remember this as the conference where I first learned (in several talks) about Brodeman's visual area 46. Just as all arrows in schematic block diagrams of the cortex point away from area 17 they seem to point towards area 46.

The quest for understanding the brain as a computational machine is not new. The father of modern philosophy and mathematics, Descartes (1596-1650) was quite interested in the brain as machine and may well be considered to be the father of computational vision. In the last paragraph of his "Treatise of Man"<sup>1</sup> (Descartes, 1664) he wrote:

"I desire you to consider, further, that all the functions that I have attributed to this machine, such as ... waking and sleeping; the reception by the external sense organs of light, sounds, smells, tastes, heat, and all other such qualities; the imprinting of the ideas of these qualities in the organ of common sense and imagination; the retention or imprint of these ideas in the memory; the internal movements of the appetites and passions; and finally, the external movements of all the members that so properly follow both the actions of objects presented to the senses and the

passions and impressions which are entailed in the memory--I desire you to consider, I say, that these functions imitate those of a real man as perfectly as possible and that they follow naturally in this machine entirely from the disposition of the organs--no more nor less than do the movements of a clock or other automaton, from the arrangements of its counterweights and wheels."

Descartes would probably appreciate that this conference on computational vision was directly in line with the program that he laid out in the above quotation. He would have in particular enjoyed hearing about area 46, a new candidate for the assimilation role that he once ascribed to the pineal gland.

This conference gave many dramatic examples of the progress being made in the Neurobiology half of the conference title. But what about progress in the first half of the title: Computational Vision. Progress has been slow in this field if judged by the inflated dreams following David Marr's inspired pioneering work<sup>2</sup>. It would have been slower yet if the explorers in these fields did not have neurobiological systems as a guide.

I suspect that progress in computer vision has been slow because the problems being worked on are too difficult. Marr set the goal of starting with a natural scene, segmenting it into isolated objects and recovering the three-dimensional shape of these objects from a single snapshot. This was a bold goal. It moved the field of vision away from using simple stimuli made of a few points, lines and gratings to real world stimuli. That may have been good since vision models should be able to be applied to the real world. The problem is that this giant leap into asking complex questions of real world stimuli was attempted before the field was walking confidently. Vision research hadn't yet gotten its feet wet dealing with even the simplest questions related to real world images, and already researchers were worrying about complex problems of segmentation, 3-d reconstruction, identifying faces and distinguishing dogs from cats. Before tackling the most difficult problems, one should first start with simpler puzzles, as discussed next.

## 2. DEVELOPMENT OF A FIDELITY METRIC: A challenge to computational vision.

An ideal short-term goal for computational vision is the development of a fidelity metric for measuring whether two real world images are perceptually identical. This

modest, achievable goal not only can bring the satisfaction of success, it can also make vision research relevant to the outside world. The development of a fidelity metric is a research area that has been seriously neglected by computational vision researchers.

In order to appreciate the need for a high quality fidelity metric one must first appreciate that we are now in the middle of a revolution in which analog images and image sequences are being abandoned in favor of digital. This transition would be surprising to an old-timer who would argue that one can concentrate information more compactly in a multilevel analog signal than in a binary signal. What the old-timer didn't realize are two important facts: 1) real-world images are tremendously redundant and 2) many image features are invisible and thus irrelevant to the human visual system. Digital images allow reduction in the redundant and irrelevant parts of the image. With this reduction, called image compression, digital has become the format of the future by a wide margin. A fidelity metric is the tool for measuring which aspects of an image or image sequence are relevant. The challenge of developing a fidelity metric is presently most actively pursued by engineers working on compression algorithms. This challenge, however, lies squarely in the domain of vision research, and the improvement of fidelity metrics should become an important enterprise for vision researchers interested in computational vision.

Development of a fidelity metric is not a new enterprise. A fidelity metric is nothing other than the calculation of the  $d'$  distance between two images, where  $d'$  is a signal detection theory concept specifying the perceptual signal-to-noise difference between the two images. Signal detection theory and many vision models are dealing with little pieces of a comprehensive fidelity metric. A number of examples will be given in the rest of this paper. The problem is that progress in this area of research has been slow. We do not yet have a general model for calculating the discriminability of two simple images. We are even further from predicting the discriminability of two real-world images and image sequences -- the real task of a useful fidelity metric. There is much to be done and the Test-Pedestal approach, the theme of this paper, offers a useful tool.

### 3. THE TEST-PEDESTAL APPROACH.

The Test-Pedestal approach is deceptively simple to describe. One can think of the task of discriminating image A from image B as the task of detecting the Test image  $T=A-B$  in the presence of the Pedestal, B. This Test-Pedestal approach has close connections to a fidelity metric and image compression where one must compare two images: the original image, A, and the image that has been compressed and decompressed, B. The fidelity metric output is the discriminability of the two images in  $d'$  perceptual units. In this section we will argue that the Test-Pedestal approach offers a powerful framework for the

development of a discrimination (fidelity) metric. A number of examples will now be offered to clarify the Test-Pedestal approach. We begin with the challenge of predicting vernier acuity.

#### 3.1 Predicting edge vernier acuity.

For many years vernier acuity was thought to be mysterious since vernier thresholds of 3 sec of arc were 10 times smaller than resolution thresholds of 30 sec. In terms of the Test-Pedestal approach the mystery is removed. Klein, Casson & Carney<sup>3</sup> pointed out that edge vernier acuity can be thought of as a line added to half the edge as shown in the left panel of Figure 1. Similarly, line vernier acuity can be decomposed into a line pedestal and a dipole test as shown in the right panel.

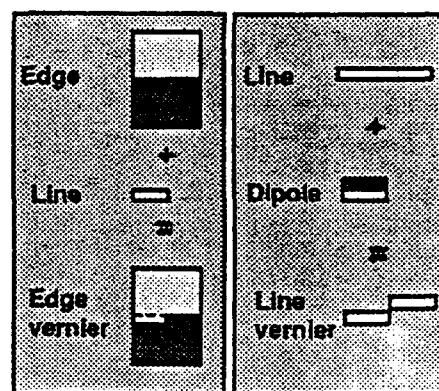


Figure 1

The idea is that instead of displaying vernier thresholds as a displacement in seconds of arc, thresholds could be presented in terms of the strength of the line that was used to produce the edge vernier offset. Figure 2 shows data for edge vernier acuity for two observers. In the top panel the vernier thresholds are plotted in min. The horizontal axis is the strength of the edge pedestal (edge strength,  $\Delta L/L$ , is twice the Michelson contrast since  $\Delta L$  is the luminance change across the edge, and  $L$  is the average luminance). In the lower panel the same data are replotted using a different vertical scale. The vertical axis is the vernier threshold in line threshold units (%min), where line strength is the product of the line contrast (%) times the line width (min). Line contrast is defined the same as edge contrast. The data points connected by a dashed line at the left of each curve are the line detection thresholds on a uniform field. They are placed horizontally at the observer's edge detection threshold. It is seen that the line detection thresholds do a very good job of predicting the vernier thresholds for low contrast edges. At higher edge pedestal contrasts the test thresholds gradually rise. That is the essence of the Test-Pedestal approach. By simply changing the units with which thresholds are plotted (from min to %min) one gains insight into why thresholds are at the level they are. In going from the top panel to the lower panel the ordinate (min) is multiplied by the abscissa (%) to obtain the new ordinate (%min). The

reverse operation can be carried out on the detection data at the left of the lower panel. These points are shown unconnected by lines on the left of the upper panel. As discussed by Klein, Casson & Carney<sup>3</sup> the procedure for calculating these points is exactly the procedure for calculating Ricco's summation width. Ricco's summation zones for these two observers are seen to be 1 and 2 min.

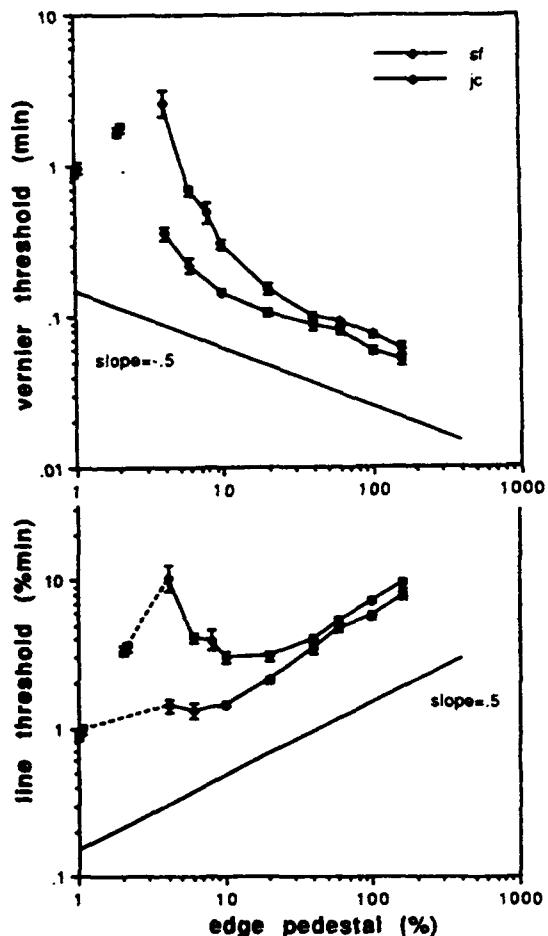


Figure 2

The bottom panel allows one to see whether the low contrast thresholds extrapolate to the uniform field detection threshold for the same test pattern (the dotted lines). One can do a similar extrapolation in the top panel. There, the low contrast vernier offset in min is expected to extrapolate to the Ricco summation extent—the leftmost data points. This is a surprising and beautiful prediction of the Test-Pedestal approach.

### 3.2 Model-free modeling.

When I presented this Test-Pedestal approach to the Asilomar conference I called it "model-free modeling". By that term I had in mind the ability to make predictions without invoking mathematical machinery and numerous modeling assumptions. I pointed out that previous predictions of hyperacuity thresholds<sup>4,5</sup> involved many

assumptions about the properties of the underlying mechanisms. The modelers making these predictions argue that there aren't many assumptions since the mechanism sensitivities and bandwidths were determined in separate experiments (using sinusoids). However, I always worry about hidden assumptions. For example, maybe the modeler explored several combination rules before finding an adequate fit to that particular data. A different combination rule might do a better job on different data. Klein<sup>6</sup> discusses other examples of hidden assumptions. The Test-Pedestal approach, on the other hand, connects edge vernier acuity directly to the line detection threshold without interruption by mathematical assumptions. No matter what approach one uses one must somehow have a method for calibrating the sensitivity of the visual system (assessing its signal to noise ratio). Previous approaches for predicting vernier acuity were excessively model dependent because sensitivity was based on the visibility of sinusoids (the CSF). Or an ideal observer approach based on photon statistics<sup>7</sup> was used. It is so much better to calibrate the visual system's sensitivity using a detection task directly related to the specific discrimination task of interest. In the example of Section 3.1 that would be line detection.

Following my talk at Asilomar there were questions about whether the Test-Pedestal approach was truly model-free modeling. One might argue that it is an assumption to say that the vernier threshold should directly extrapolate to the line detection threshold. I guess what I meant was that the data shown in Fig. 2 provide an adequate explanation of vernier thresholds at low pedestal contrast. Incidentally, similar results are found for vernier acuity using lines<sup>3</sup> and sinusoids<sup>8</sup>. The demonstration that low contrast edge vernier acuity is well predicted by the line detection threshold does not eliminate the need for modeling. One must still develop a theory for the visibility of lines in terms of the underlying neurobiological mechanisms, and one must develop a model for how the line threshold increases as the edge pedestal increases - the topic of masking. The difficult challenge of developing a model of masking is not to be underestimated. It is the biggest challenge facing the development of a fidelity metric. The rest of this article is concerned with masking. Before getting to these difficult model-ridden problems it is worth pausing for five minutes and being happy that at least abutting vernier acuity for low contrast stimuli is no longer one of the bothersome mysteries confronting vision modelers.

In a sense what the Test-Pedestal approach says is that one should always look at discrimination data as a function of the pedestal strength. The plot of the test threshold vs. pedestal strength is often called a threshold vs. contrast (tvc) curve. It should maybe be called a tvs curve (threshold vs strength) since the abscissa for line vernier would have line strength units of %min, but for historical reasons we will stick with tvc. In any case, pedestal strength is linearly related to pedestal contrast.

The tvc curve has the advantage that in addition to plotting the discrimination data (like vernier thresholds) one can also plot the detection threshold of the test pattern on a uniform field (like the line detection threshold). This detection threshold is the expected vernier discrimination threshold when the pedestal strength is small.

How does one explain that thresholds rise as pedestal strength increases? The tvc curve in Fig. 2 increases with a slope of about .5. A straight line with slope of .5 is shown below the data for comparison. One could try to develop a model for this slope (the slope of .5 conjures up thoughts of Poisson noise ideal observer models). Or one could continue with the model-free strategy and try to account for the tvc slope in terms of other masking data. For example, vernier acuity which involves the masking of line visibility by an edge pedestal, could be compared to line contrast discrimination which involves the masking of line visibility by a line pedestal. That is the strategy successfully followed in a set of experiments by Carney and myself<sup>9</sup>. The trouble with this approach is that the two pedestals have quite different spatial frequencies so one could worry that the masking functions need not be similar. For that reason Hu, Klein & Carney<sup>8</sup> carried out a series of experiments involving sinusoidal vernier acuity and compared them to sinusoidal contrast discrimination.

### 3.3 Sinusoidal vernier acuity-the need for optimal conditions.

The pedestal stimulus in the Hu, et al.<sup>8</sup> experiments was a sinusoidal grating that can be written as:

$$S_p(x, y) = c_p \cos(fx) \quad (1)$$

where  $c_p$  is the pedestal contrast. The pedestal plus test pattern for contrast discrimination is:

$$S_{p+t}(x, y) = (c_p + c_t) \cos(fx) \quad (2)$$

where  $c_t$  is the test contrast. The test pattern is added to half of the pedestal (for  $y>0$ ). The difference between the two patterns is the test pattern:

$$T_{jnd}(x, y) = c_t \cos(fx). \quad (3)$$

For the vernier stimulus the pedestal plus test pattern is:

$$S_{p+}(x, y) = c_p \cos(fx+\phi) \quad (4)$$

where  $\phi$  is the phase shift of one half of the grating ( $y>0$ ) due to the vernier offset. The difference between the pedestal plus test and the pedestal alone is given by:

$$\begin{aligned} T_{vern}(x, y) &= S_{p+}(x, y) - S_p(x, y) \\ &= c_t \sin(fx+\phi/2) \end{aligned} \quad (5)$$

with  $c_t = c_p 2 \sin(\phi/2)$ . (6)

This is the derivation given by Hu, et al.<sup>8</sup> for the test pattern for the vernier stimulus. It all seems quite reasonable and straightforward based on trigonometry.

There was, however, a conceptual error in the Hu, et al. derivation. The test contrast given by Eq. 6 is relevant to the task of detecting the vernier offset. In our experiments, however, the observer's task was that of discriminating a rightward offset from a leftward offset. For a discrimination task the true test pattern becomes:

$$T_{vern-disc}(x, y) = (S_{p+}(x, y) - S_{p-t}(x, y))/2 \quad (7)$$

where  $S_{p-t}(x, y)$  is a displaced sinusoid where the sign of  $\phi$  is reversed. The test pattern can be written as:

$$T_{vern-disc}(x, y) = c_t \sin(fx) \quad (8)$$

where  $c_t = c_p \sin(\phi)$ . (9)

Eq. 9 gives the correct test contrast relevant to the discrimination task, rather than Eq. 6 which was specified by Hu, et al.<sup>8</sup>. Luckily the difference between the two values is too small to have made a difference in any conclusions of that article (the corrected data in Fig. 3 can be compared to the data in the original article). We have gone into this level of detail in order to point out that when using the Test-Pedestal approach one must be careful in how one defines the test pattern.

The beauty of using sinusoidal stimuli is that just by a 90 deg change in the phase of the test pattern (compare Eqs. 2 and 8) one can switch from a vernier stimulus to a contrast discrimination stimulus. The two tasks can therefore be directly compared.

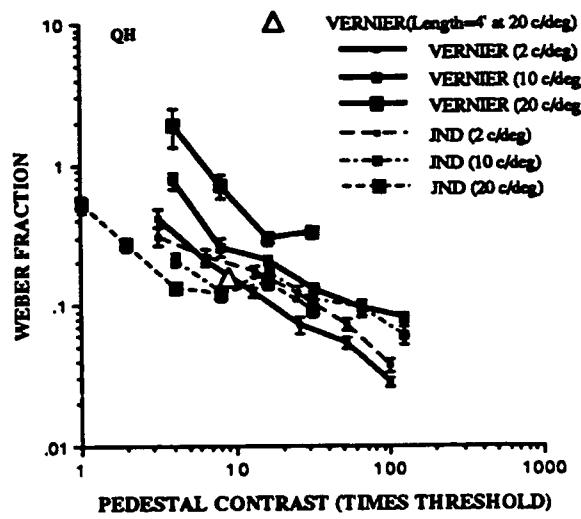


Figure 3

Fig. 3 shows the tvc curves for one observer at three spatial frequencies, 2, 10 and 20 c/deg. Experimental details and results of other observers and other spatial frequencies can be found in Hu, et al.<sup>8</sup> The open and filled symbols are for the vernier and contrast discrimination tasks respectively. The data show that except for 20 c/deg the vernier and contrast discrimination thresholds are about equal. Two differences are visible. First, the log-log slope of the vernier tvc curve is slightly steeper than the contrast discrimination slope (-.61 vs -.44). If the data had

been plotted with test contrast on the ordinate, the vernier slopes would have been shallower (.39 vs .56). This would be expected from a filter model since at low pedestal contrast the contrast discrimination task, but not the vernier task, should show facilitation. At high contrast, on the other hand, the vernier task has lower contrast thresholds because the oriented mechanisms used for the vernier task are tilted away from the edge direction and thus are less affected by high mask contrasts. The second difference is the degradation of vernier thresholds at high spatial frequency, 20 c/deg. Hu, et al.<sup>8</sup> made a major effort to understand the degradation at high spatial frequency. In the next paragraph we examine this loss to gain further insight into the range of validity of the Test-Pedestal approach.

It is not surprising that at high spatial frequency vernier is degraded more than contrast discrimination. In the contrast task a comparison between the two halves of the screen is not needed whereas the vernier task requires precise positioning of the mechanisms. At high spatial frequencies the precision of positioning becomes more critical. Furthermore, in the vernier task since the optimal mechanisms are tilted, their ends will overlap inappropriate regions of the stimulus for high spatial frequencies. In order to explore this possibility, Hu, et al.<sup>8</sup> repeated the vernier and contrast discrimination experiments with very short stimuli. For the contrast discrimination task thresholds became elevated, whereas for the vernier task thresholds decreased more than two-fold at 20 c/deg. The triangle in Fig. 3 is the datum for the short grating for the vernier task. It is seen that even at 20 c/deg the vernier thresholds are in good agreement with the contrast discrimination thresholds as long as each is measured under optimal conditions. It may not always be possible to find optimal conditions under which discrimination thresholds equal detection thresholds. If one does discover these optimal conditions then one has gained insight into receptive field properties of underlying mechanisms. If one does not find optimal conditions then one must search for other sources of noise that interfere with the discrimination. An example where other noise sources are dominant is taken up next.

#### 3.4 A fundamental position limit? Insight from strabismic amblyopes and peripheral vision.

Tvc plots of edge vernier acuity<sup>10</sup> often show that at the very highest edge contrasts the vernier thresholds reach a floor limit of about 6 sec (corresponding to 4 sec if threshold is defined at 75% correct rather than the d'=1, 84% correct). It is possible that a new noise source is present that has a fundamental spatial uncertainty of about 6 sec. Spatial limitations to position acuity would then become the limiting factor for very high contrasts.

The examples of vernier acuity given so far have been for abutting stimuli presented to the fovea of observers with normal vision. In these examples vernier acuity at low contrast extrapolates beautifully to the relevant

detection threshold (a line in Section 3.1 and a sinusoid in Section 3.3). The story is different in peripheral vision and the vision of strabismic amblyopes. In these two degraded visual systems vernier acuity is degraded from what one would expect based on detection<sup>10, 11, 12</sup>. There can be as much as 3-fold and 10-fold extra loss in peripheral vision and amblyopic vision respectively. Vernier acuity of anisometropic amblyopes, on the other hand, was found to be compatible with what is expected from their degraded detection capability<sup>11, 12</sup>. Similar results had been obtained from previous research not using this Test-Pedestal approach<sup>13, 14</sup> but in those previous studies vernier acuity was compared to sinusoid detection so one couldn't make a direct prediction for the low contrast vernier thresholds.

The straightforward interpretation of the extra degradation in the periphery and in strabismics is that in these visual systems the visual "grain" is coarser than the 6 sec limit discuss at the beginning of this section. There is a problem with this hypothesis of a spatial floor. A simple spatial floor would have a flat threshold as contrast is reduced until the Test-Pedestal line detection threshold was reached. However, we found that both in the periphery and in strabismics thresholds degrade at low contrast<sup>10, 11</sup>. In fact the slope is the same as that found for normal foveal vision. These complexities are discussed (but not fully resolved) by Levi & Klein<sup>11</sup>.

#### 3.5 Vernier acuity with gaps. Sources of noise dependent on separation and eccentricity.

There is a dramatic case in which the Test-Pedestal approach fails: vernier acuity with a large gap. The presence of a large gap, of course degrades vernier thresholds but doesn't severely affect contrast discrimination, which stays pretty much at a 10% Weber fraction independent of gap. Hu, et al.<sup>8</sup> measured both vernier and contrast thresholds for different gaps, including an "infinite gap" for contrast discrimination in which the reference is not shown simultaneously. The gap effect is related to the dramatic falloff of vernier acuity with eccentricity. Levi & Klein<sup>15</sup> showed the two cases are connected since a large gap places the stimulus in the periphery. This dramatic "violation" of the Test-Pedestal framework is easy to understand in terms of four different regimes that place different limits on vernier acuity.

The four regimes for vernier acuity depend on the size of the gap between the relevant features. Consider first vernier acuity for dots (either 2 or 3 dots). Levi & Klein<sup>15</sup> discuss three of these regimes: for a very small gap (<1.5 min) in the resolution regime the dots become blurred together and vernier acuity becomes degraded. For wide separations (>=30 min) we have the local sign regime where the limited spatial resolution of peripheral vision becomes the main limit to position acuity. In between the very large and the very small separations is the filter regime. The filter regime occurs when the feature

separation is much smaller than the feature eccentricity. The local sign regime occurs when the separation is about the same extent as the eccentricity (for example, a wide separation 2-dot or 3-dot vernier task centered on the fovea-unless the separation is smaller than about 30 min). In the filter regime vernier thresholds are based on the output of an orientation tuned filter that spans the critical features of the stimulus. In the filter regime with high contrast stimuli, vernier thresholds are approximately<sup>15</sup>

$$Th = (sep + .1)/30 \quad (10)$$

where sep is the separation of the features (the gap in deg). These experiments were carried out on an isoeccentric arc to remove the effects of eccentricity. The vernier offset was in the radial direction which is the direction in which vernier thresholds are about a factor of two poorer than offset thresholds in the tangential direction<sup>16</sup>. Klein & Levi<sup>17</sup> found thresholds up to four-fold lower than Eq. 10 using linearly arranged 3-dot stimuli centered on the fovea.

$$Th = (sep + .1)/100 \quad (11)$$

Thus for a separation of 6 min the thresholds are approximately 7 sec (corresponding to about 5 sec at a 75% correct criterion). The lower thresholds in Eq. 11<sup>17</sup> as compared to Eq. 10<sup>15</sup> are undoubtedly due to having the central dot in the fovea rather than in the periphery; and having the offset in the tangential rather than radial direction. In this filter regime vernier acuity is limited by the orientation tuning of filters. For larger gaps, larger filters are used. These filters can signal orientation to about .6 deg, thereby accounting for thresholds of about 1/100 of the separation. Where does the factor of 1/100 come from? It is likely based on the orientation tuning of the underlying mechanisms. The orientation tuning is about 10 deg<sup>18</sup> (half-bandwidth at half height) corresponding to 1/6 of the separation. The 1/100 orientation discrimination is about 6% of the bandwidth. This number of 6% for discrimination tasks is close to the 10% Weber fraction that is found in contrast discrimination. A similar factor relates spatial frequency discrimination to spatial frequency bandwidths. This argument is quite crude (that is why it is appearing in a conference proceeding rather than in JOSA). Appendix 5 of Klein & Levi<sup>4</sup> has a more formal argument of this sort for spatial frequency discrimination.

We advertised four regimes, but have only mentioned three so far. The fourth is the Test-Pedestal regime that places a floor on the optimal vernier acuity. The .1 in Eqs. 10 and 11 is present to indicate a transition between different sources of noise limiting vernier acuity. For separations larger than about 6 min ( $sep \gg .1$  deg) the orientation tuning of filters limits thresholds. For separations that are smaller than 6 min (but larger than the resolution separation of about 1 min) the limitation is no longer orientation tuning but rather the Test-Pedestal limit that is the theme of this paper. Thus for the line vernier task, the limit would be the visibility of a dipole<sup>3</sup>. The data shown in Figs. 2 and 3 were for abutting vernier

tasks so they were in the Test-Pedestal regime. This is the regime where one obtains the best thresholds. This is not to say that the properties of the filter mechanisms aren't relevant, it is just that the orientation tuning isn't the only consideration. As was discussed in Section 3.3 the properties of the filters can be seen in the subtle differences between vernier and contrast thresholds. When the gap becomes greater than about 6 min, then the filter orientation tuning becomes the limiting factor for the vernier task. Since orientation tuning isn't relevant to contrast discrimination, these two tasks begin to differ in their properties for large gaps.

In the local sign regime where the separations are comparable to the eccentricity, thresholds are limited by an intrinsic uncertainty in position that increases with eccentricity. This uncertainty places strong limits on the position acuities (like vernier acuity), without placing limits on contrast discrimination. Vernier acuity is expected to fall off according to

$$Th = .01(E+E_2) \quad (12)$$

where E is the eccentricity of the most distant feature of the vernier task and  $E_2$  is between .6 deg and 1 deg (depending on stimulus orientation, and visual field meridian) for vernier acuity. Thus at an eccentricity of 10 deg the vernier thresholds are about  $.01*(10+1)\text{deg}=6.6$  min, more than 60 times the foveal value. The local sign limitation is well understood in terms of the cortical magnification factor. Peripheral vision has greater position uncertainty than foveal vision. Thus position tasks are degraded. Contrast discrimination, on the other hand, does not require spatial comparisons so it is unattenuated in the periphery. We have gone into this detailed discussion of cases in which vernier acuity can differ from contrast discrimination in order to put the Test-Pedestal approach in its proper context. One must be somewhat careful with the claim that discrimination can be directly related to a detection or contrast discrimination task. One must be careful to avoid other sources of noise that can severely limit the discrimination.

We have examined a number of cases in which discrimination can be worse than detection. Can it be better? One is probably safe in asserting that a discrimination task can never be more than 3 times better than the comparable detection task. The factor of three is present because facilitation can sometimes reduce detection thresholds by that factor. Next we will consider a one-dimensional example in which the discrimination task does indeed exhibit facilitation.

### 3.6 Discriminating edge blur and square wave-sinusoid discrimination. A simple model.

Campbell & Robson's<sup>19</sup> paper had a strong influence in getting the "Fourier analysis of vision" bandwagon going. One of its claims was actually a beautiful example of the Test-Pedestal framework. Campbell & Robson's

data showed that a square wave could be discriminated from a sine wave of the same fundamental spatial frequency when the third harmonic was at its independent threshold. This is an example of the Test-Pedestal approach with the third harmonic as the test pattern (the third harmonic is the main component of the difference between a square wave and a sinusoid) and the fundamental as the pedestal. The importance of their finding goes beyond the task of discriminating two grating profiles. This particular discrimination task is nothing other than a general blur discrimination task. The observer makes his judgment on the sharpness of the "edges" of the grating. Blur discrimination has strong relevance to many visual tasks including accommodation and fidelity metrics.

An interesting sidelight to the Campbell & Robson paper is the attempt by Stromeier & Klein<sup>20</sup> to replicate their results. Rather than discriminating a square wave from a sinusoidal grating we simplified the stimulus (in Fourier space) to the task of detecting a 9 c/deg sinusoidal grating (third harmonic) when added to a 3 c/deg static pedestal (the fundamental). Contrary to the Campbell & Robson<sup>19</sup> result we found facilitation for a wide range of pedestal contrasts (Campbell & Robson didn't report facilitation possibly because they used the method of adjustments). We also did it with a 9 c/deg pedestal so the task was contrast discrimination. We again found facilitation (simultaneously found by Nachmias and Sansbury<sup>21</sup>). We did the same experiment with a 1.8 c/deg pedestal (1st & 5th harmonic) and found neither facilitation nor threshold elevation of the 5th harmonic. A theory explaining these results was developed in terms of medium bandwidth mechanisms and an accelerating transducer function (Stromeier & Klein<sup>20</sup>). Variants of this approach were later pursued by Legge & Foley<sup>22</sup> and by Wilson and collaborators<sup>23</sup> with good success.

The Stromeier & Klein<sup>20</sup> model has been a prototype for many successive models so it would be nice to see how it works in detail. One interesting feature of the model is that a "continuum" of mechanism sizes are used. Then a search is done for the optimal mechanism. In order to be fully clear about the model the following Matlab code presents it in full. For simplicity we take the fundamental and third harmonic to be at 1 and 3 c/deg.

```

1 function transducer=trans(contrast)
2 W=2; transducer=log(1+contrast.^3/W)/log(1+1/W);
-----
3 function resp=cauchy(f, n)
4 resp=(f.*exp(-f+1)).^n;
-----
%The following is the main routine
5 cont_ped=0:50; freq_mech=2:.02:3.5; n_cauchy=6;
6 eff_ped=cont_ped'*cauchy(1./freq_mech, n_cauchy);
7 eff_test=ones(51,1)*cauchy(3./freq_mech, n_cauchy);
8 diff=trans(eff_ped+eff_test)-trans(eff_ped);
9 [maxdiff i_order] = max(diff);
10 diff_all=[maxdiff;diff(:,[6 21 36 51 66])];

```

```

11 subplot(211); plot(cont_ped,diff_all)
12 ylabel('detectability (d prime)')
13 subplot(212); plot(cont_ped, 2+(i_order-1)/50);
14 ylabel('optimal mechanism'); xlabel('pedestal contrast')

```

Line 1: The transducer function is defined. The input is contrast (in threshold units) and the output is d'.

Line 2: The transducer function is taken from Klein & Levi<sup>4</sup>. In that paper the exponent was 2 whereas here an exponent of 3 is used because of the steep acceleration (Stromeier & Klein<sup>20</sup> used an exponent of 4). The Weber parameter, W, is taken to be W=.5 just as in Klein & Levi<sup>4</sup>.

Line 3: The Cauchy function is defined (see Klein & Levi<sup>4</sup>). The first input is the spatial frequency where the peak spatial frequency is taken as unity. The second input is the Cauchy exponent that specifies the mechanism bandwidth.

Line 4: This is the Cauchy formula normalized so that it has a peak value of 1 at f=1. For the present calculations medium bandwidth mechanisms with n=6 are used.

Line 5: Defines the pedestal contrast to go from 0 to 50 contrast threshold units and defines the peak spatial frequency of the model's mechanisms to go from 2 c/deg to 3.5 c/deg in steps of .02 c/deg. This range of mechanisms is chosen to encompass the full range of mechanisms relevant to the task. The Cauchy index is chosen to be 6 corresponding to a medium bandwidth mechanism.

Line 6: Defines effective pedestal contrast to be the pedestal contrast times the mechanism tuning function's sensitivity to 1 c/deg. This is done for each pedestal contrast and each mechanism. The ratio 1/freq\_mech occurs because 1 c/deg is the frequency of the pedestal pattern and freq\_mech is the peak frequency of the mechanism.

Line 7: Similar to line 6 except for two items: 1) The effective test contrast is defined to be a unity test contrast (defined by the "ones function") times the mechanism tuning function's sensitivity to 3 c/deg. The ratio 3/freq\_mech occurs for a similar reason as the comparable factor in line 6, except now 3 c/deg is the test frequency.

Line 8: The differential response is calculated. This is the d' for the test plus pedestal minus the d' for the pedestal alone. The transducer function gives the d' as a function of contrast.

Line 9: The optimal mechanism is determined by taking the maximum differential response at each pedestal contrast. Matlab's max function takes a matrix as input and outputs two row vectors: maxdiff and i\_order. . Maxdiff consists of the maximum value in each column, and i\_order is the index of the maximum value. Line 13 converts this index to the peak frequency of the optimal mechanism.

Line 10: An array of differential responses is created. The first row is the maximum differential response at a given pedestal contrast. The maximum is taken over all spatial frequency mechanisms. The second row is the differential response of the sixth mechanism, corresponding to a peak spatial frequency of 2.1 c/deg.

In succeeding rows the peak spatial frequency increases by .3 c/deg for each row. This sparse sampling of mechanisms simplifies the plot.

Line 11: The subplot command allows two plots in each figure. The plot command plots all the differential responses as a function of the pedestal contrast. The different lines are: upper envelope-solid, peak frequency (c/deg) of 2.1-dashed, 2.4-dotted, 2.7-dot-dashed, 3.0-solid, 3.3 dashed.

Line 12: The ordinate for the top plot is labeled.

Line 13: The lower panel is generated. Based on the its definition in line 5 i\_order values of 1 and 51 correspond to peak frequencies of 2 and 3 c/deg. Other values are linearly interpolated.

Line 14: Labels for the bottom panel.

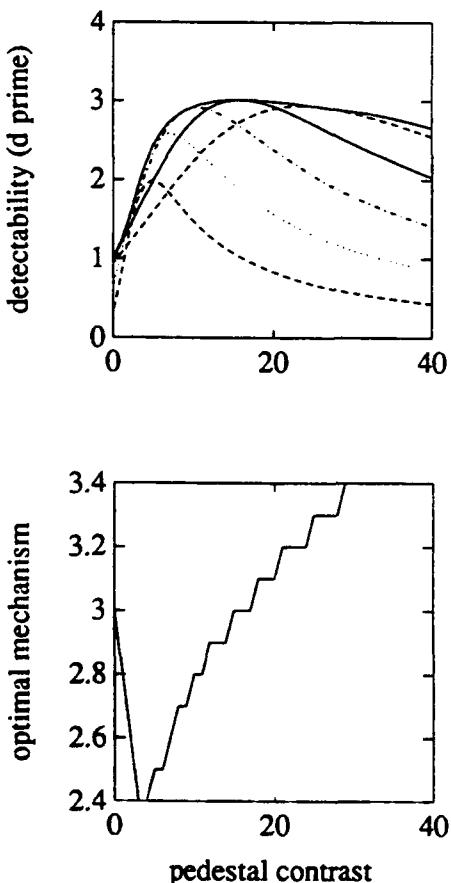


Figure 4

The upper panel shows the  $d'$  responses of several mechanisms whose peak spatial frequencies range from 2.1 to 3.4 c/deg. The upper solid line is the envelope of the response of the individual mechanisms. It represents the observer's response assuming a peak detection model. The lower panel shows which mechanism produces the largest  $d'$  (a smoother curve would have been produced by a finer sampling of the mechanisms). The dotted line is for the mechanism whose peak frequency is 2.4 c/deg. From both the upper and lower panels it is seen that this mechanism provides the best signal to noise when the pedestal is between 3 and 4 contrast units. This

mechanism is optimally situated so that the 1 c/deg pedestal brings the mechanism to the facilitation region where it is highly sensitive to the test. As the pedestal contrast is raised further this mechanism begins to saturate and a higher spatial frequency mechanism becomes optimal. The lower solid line corresponds to the 3 c/deg mechanism. It becomes the optimal mechanism twice: First, at zero pedestal contrast when the test pattern is presented on a black background. Of course the 3 c/deg mechanism is optimal here since that is the frequency of the test pattern). Second, from 15-17 pedestal contrast units, as the optimal mechanism moves away from the pedestal. At yet higher pedestal contrasts the optimal mechanism has a spatial frequency higher than 3.0 c/deg in order to avoid the masking effects of the pedestal. These higher frequency mechanisms maintain their response to the test while their response to the pedestal falls rapidly.

We have gone into such detail for this task of detecting a third harmonic in the presence of a fundamental for two reasons: First, we wanted to emphasize the subtlety that the optimal mechanism need not be the mechanism that detects the test pattern on a blank background. Second, as mentioned earlier this task is directly related to the task of detecting blur of a square wave grating. The task of detecting edge blur is central to the enterprise of developing a good fidelity metric, an important motivation for this paper.

We have pursued this question of detecting edge blur for single edges as well as for gratings. We measured the visibility of edge blur as a function of edge contrast using the Test-Pedestal framework<sup>9, 24</sup>. The difference between a sharp edge and a blurred edge (with a threshold amount of blur) is a dipole. We found that edge blur can be discriminated below the threshold for detecting a dipole on a uniform field. For a wide range of pedestal contrasts (edge contrast) the blur threshold (in dipole units) is about the same as dipole contrast discrimination at the bottom of the dipper function. The edge pedestal is facilitating dipole detection, similar to our finding with the first plus third harmonic experiment.

### 3.7 Monopolar and bipolar cues and mechanisms. Dependence on how threshold is defined

Here is an interesting problem. In the Test-Pedestal approach one compares discrimination thresholds to detection thresholds. The problem is that the transducer function relating  $d'$  to stimulus strength tends to be different for detection and discrimination tasks.

In order to clarify how the transducer shape affects threshold we must be precise about how thresholds are defined. The connection between  $d'$ , stimulus contrast,  $c$ , and threshold,  $th$ , is given by:

$$d' = d_t (c/th)^n \quad (13)$$

where  $d_t$  is the  $d'$  level at which threshold is defined ( $d_t = 1$  and .68 for threshold defined at 84% and 75% correct respectively),  $n$  is the transducer exponent ( $n=1$  or 2 for discrimination or detection respectively). Eq. 6 was written so that when the contrast is at threshold ( $c=th$ ) then  $d'=d_t$ .

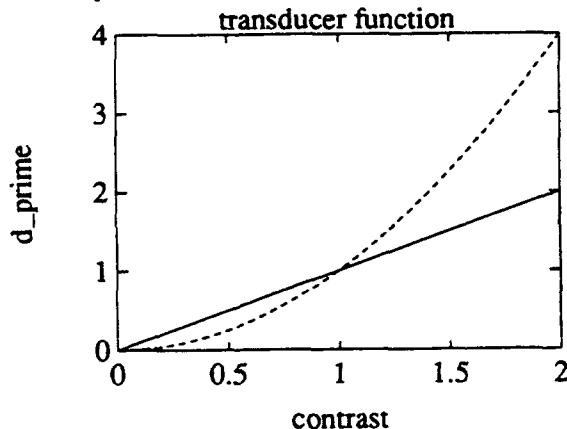


Figure 5

This figure was produced by the following Matlab code:  
 $x = 0:1:2; y(1,:) = x; y(2,:) = x.^2; plot(x,y)$   
 $xlabel('contrast'); ylabel('d_prime'); title('transducer function')$

Fig. 5 shows two transducer functions, with  $n=1$  and  $n=2$ . They would imply the same threshold (detection and discrimination thresholds are equal) if threshold were defined to be at  $d_t = 1$ . However, if  $d_t < 1$  then the discrimination threshold is lower than the detection threshold. The reverse order is obtained if  $d_t > 1$ . Suppose for example threshold is defined to be at  $d'=2$ . Then for discrimination the threshold strength would double but for detection it would increase by  $\sqrt{2}$ . Thus whereas the two thresholds would be equal for  $d_t=1$  they differ by  $\sqrt{2}$  for  $d_t=2$ . This dependence of threshold on one's choice of detection criterion affects the interpretation of the test-peDESTAL results, so it shall be examined here.

The rest of this section explores why detection and discrimination are expected to have different shaped transducer functions. Detection tasks tend to have accelerated transducer functions<sup>20,21,25</sup> whereas discrimination tasks usually have a linear transducer function. An insight into the basis for this difference is the monopolar-bipolar distinction discussed by Klein<sup>26</sup>. Consider edge vernier acuity. There are two types of cues that can be used for the vernier judgment.

1) **Bipolar cue.** The vernier offset provides an orientation cue that is bipolar. A cue is bipolar if a negative cue is perceived as being in the opposite direction as a positive cue. For a bipolar cue if a rightward offset can be discriminated from a blank (zero offset) with  $d'=1$  (84% correct for the stimulus with an offset and 50% correct for the blank), then a leftward offset of the same amount should be discriminable from a rightward offset with  $d'=2$  (84% correct on each stimulus).

2) **Monopolar cue.** The vernier offset can potentially also have a monopolar cue based on detecting a break in the line without information about the direction of the break. The monopolar mechanism produces a positive response for both positive and negative offsets, whereas the bipolar mechanism produces a negative signal for a negative offset. Based on the monopolar cue the discrimination of a rightward from a leftward offset would have  $d'=0$ .

How do the transducer functions differ for monopolar and bipolar mechanisms? Monopolar mechanisms tend to have a quadratic transducer function,  $d'=s^2$ , (the same response to positive and negative stimuli). Bipolar mechanisms tend to have a linear transducer function,  $d'=s$ , (antisymmetric response to positive and negative stimuli). It is conceivable for the mechanisms to have a different behavior. The monopolar mechanisms could behave linearly, but this would require a full-wave rectification behavior,  $d'=|s|$ . This is unlikely since it involves a singularity at  $s=0$ , and nature tends to avoid singularities.

A bipolar mechanism could deviate from a linear response, i.e. it could have a cubic response, but that is unlikely. It is especially unlikely for the many discrimination tasks in which the reference stimulus is not special. For example, in contrast discrimination or 2-dot vernier acuity the reference stimulus is a pedestal that is not qualitatively different from the positive or negative stimuli. In that case one can make a Taylor's series expansion of the transducer function,  $T(\text{contrast})$  around the reference point.

$$d' = T(p+t) - T(p) \\ = T'(p) t + \text{terms of order } t^2 \quad (14)$$

The  $d'$  is equal to the difference between the transducer response to the test plus pedestal,  $p+t$ , minus the response to the pedestal alone. If the pedestal reference is not a special stimulus then the first derivative in Eq. 14 will not vanish and  $d'$  is seen to be linearly. If the pedestal is special then the first derivative could vanish. Klein<sup>26</sup> gives several examples where the pedestal is special by being at a natural zero of the stimulus. For example, a blank field is a natural zero, such that the first term of Eq. 14 is able to vanish. The leading term would then be quadratic in stimulus contrast. This would make increments and decrements look the same.

We now come back to the question posed at the beginning of this section: How is the monopolar-bipolar distinction related to the detection-discrimination distinction. To first order, detection is a monopolar task since in many cases (such as detecting a high spatial frequency grating) one can not discriminate the positive from the negative stimulus when at threshold. Similarly, to first order, discrimination is a bipolar task, since the presence of the pedestal allows the polarity of the test to be discriminable. The clearest example is contrast

discrimination where an increment and decrement of the pedestal are clearly in opposite directions. One must be careful, however, because there are many exceptions to this connection between monopolar with detection and bipolar with discrimination. Consider the following examples where detection may be bipolar:

- 1) The detection of a low spatial frequency grating may have a bipolar component. A .5 c/deg grating has a low enough spatial frequency that the observer would know where to look for the bright and dark bars (we are assuming phase is not randomized trial to trial). Thus the task might be done by a luminance discrimination: is the luminance at fixation higher or lower than the luminance of the surround- a bipolar judgment.
- 2) Detection of an edge. Is the right half of the field brighter or dimmer than the left half. Again, this is a bipolar judgment.
- 3) Detection of a light or dark line. Again, one might use a bipolar judgment here. A detailed analysis of line polarity identification and detection was done by Klein<sup>26</sup>. There was insufficient data to measure the transducer function shape, but since the polarity could be identified near detection threshold there was evidence for a sensitive bipolar mechanism.

### 3.8 Bisection hyperacuity. Predicting the Guinness world record for visual acuity

Starting in 1991, the Guinness Book of World Records<sup>27</sup> includes the following entry for visual acuity.

**Highest visual acuity.** The human eye is capable of judging relative position with remarkable accuracy, reaching limits of between 3 and 5 sec of arc.

In April 1984 Dr. Dennis M. Levi of the College of Optometry, University of Houston, TX repeatedly identified the position of a thin white line within 0.85 sec of arc. This is equivalent to a displacement of some 1/4 inch at a distance of 1 mile.

The full experimental details of this Guinness experiment and related experiments are to be found in Klein & Levi<sup>4</sup>. Let me here only focus on the conditions of the Guinness experiment. Five very bright horizontal lines were displayed on a dark background. The middle line, called the test line, was surrounded by a pair of reference lines, which were in turn surrounded by a pair of flanking lines. The test-to-reference separation was 1.3 min (when the test was centered) and the reference-to-flank separation was 1.2 min. On a given trial during an experimental run the test line was shifted by a small amount either up or down. The observer's task was to identify the direction of the shift. We found that both observers had thresholds of less than 1 sec. Observer D.L. repeated this experiment 8 times. His average threshold was  $.85 \pm .04$  sec. We are defining threshold to be 75% correct rather than 84% correct. That is, 75% of the time an observer could

correctly identify an upward displacement of .85 sec from a downward displacement of .85 sec. This level of detection corresponds to a d' of .68 of discriminating a .85 sec offset from a zero offset, or a d' of 1.36 of discriminating offsets of +.85 and -.85 sec. Now we ask how can one predict these results from the properties of the underlying mechanisms.

Most of the Klein & Levi<sup>4</sup> paper is devoted to developing the viewpoint model. The viewpoint model is an extension of the models we developed almost 20 years ago<sup>20, 28</sup>. The pros and cons of these models were discussed by Klein<sup>6</sup>. Here we merely note that in any filter model there are many assumptions about filter sensitivity, filter non-linearities, and rules for combining filter outputs. One might argue that these assumptions are not associated with free parameters since the parameters of the model can be determined in prior studies (such as masking studies). However, since experimental conditions between experiments are usually different, one must still make judicious guesses about how the parameters of one experiment apply to a different experiment.

Instead of applying a detailed filter model, we have taken two approaches to accounting for the bisection data. First, Klein & Levi<sup>4</sup> presented five Appendices with simplified models for getting a ballpark estimate of optimal bisection thresholds. These methods show how a simplified Fourier approach can be combined with a spatial approach to provide insight into psychophysical tasks. I continue to find these five Appendices as valuable reminders for how to think about modeling. Second, is the Test-Pedestal approach which we now examine. A different version of this analysis, together with a more formal introduction to the Test-Pedestal approach and its relationship to Geisler's Ideal Observer model was presented by Klein<sup>6</sup>.

In order to make it easier to understand the workings of the Test-Pedestal approach it is useful to first consider a very similar experiment in which the middle three lines are not idealized as infinitely thin (an impossible stimulus actually) but rather have a width of s=1.3 min. This rectangular blurring operation makes the three lines just barely touch when the central line is exactly bisecting the two reference lines. This touching occurs because the rectangular blur has the identical width as the separation between test and reference lines. For the present argument it doesn't matter whether one also blurs the two flanking lines.

Now suppose the central line is shifted upward by a small amount,  $\delta$ . At points .65 min below and above the midpoint there will be thin lines of width  $\delta$  that are black and white respectively. The white line's luminance will be twice that of the local average luminance. Thus the black and white lines have contrasts of -100% and +100%. This adjacent black-white combination is called a dipole. Thus the shifted middle line can be replaced by an unshifted line plus the test dipole.

The strength of a dipole, called the dipole moment, is given by the product of the strength of each line, called the line moment, times the line separation. The strength of a line is the product of the line contrast times the line width. In our case the strength of each line is  $\pm 100\delta$  %min, and the separation of the two lines is 1.3 min. Thus the dipole strength is  $\pm 130\delta$  %min<sup>2</sup>. For the Guinness record offset of  $\delta = .85$  sec the corresponding dipole strength is 1.84 %min<sup>2</sup> or 2.7 %min<sup>2</sup> if the d'=1 criterion had been used for bisection threshold. Now for the punch line. It turns out that the detection threshold for a dipole on a uniform field is about 2 %min<sup>2</sup><sup>29</sup>. Thus the Guinness record bisection threshold can be understood simply in terms of the visual system's sensitivity as measured by detection threshold. The suprathreshold pedestal did not mask the visibility of the dipole. The role of the flanking lines can now be understood in terms of enlarging the uniform field on which the dipole is detected<sup>4</sup>. For 3-line bisection the stimulus width is less than 3 min, which presumably is insufficient for optimal sensitivity of the dipole detection mechanism. By adding the two flanking lines the background is more than 5 min in width, thereby providing an adequately wide platform. This explanation of the optimum bisection threshold is much more direct than the assumption-ridden modeling provided by Klein & Levi<sup>4</sup>. However, we still need the filter model<sup>4</sup> to account for the behavior of the bisection threshold as the separation between the five lines are modified. The filter model is needed to account for why thresholds are 1/60 of the separation, similar to what was discussed in Section 3.5 in connection with vernier acuity.

### 3.9 Motion discrimination. More problems for the Test-Pedestal approach.

As a final example of the Test-Pedestal approach we consider motion detection and discrimination. One might suspect that motion would present difficulties for the Test-Pedestal approach. This is because there is a belief among vision researchers that the motion system saturates at much lower contrast than the pattern system. Some invoke the magno (motion) - parvo (pattern) distinction. This belief is based on the Nakayama-Silverman experiment on the detection of a displacement of a sinusoidal grating. We have just written a pair of papers that claim the opposite. We now give a very brief summary of these papers:

- 1) Beard, Klein & Carney<sup>30</sup> used a static sinusoidal mask and a counterphase test of the same spatial frequency presented either in-phase or in quadrature phase with the test. The quadrature phase stimulus appeared as a grating oscillating back and forth in spatial position and the in-phase stimulus appeared as a grating oscillating in contrast. The finding was that over a wide range of spatial frequencies, temporal frequencies and pedestal contrasts, the motion stimulus had the same detection threshold as the contrast stimulus. In addition the in-phase and out-of-phase stimuli could be discriminated at the detection

threshold. Here we merely want to emphasize that these results can be thought of as a vindication of the Test-Pedestal approach. In fact, these experiments are quite similar to the sinewave vernier experiments of Hu, et al.<sup>8</sup> discussed in Section 3.3. Those experiments were a zero Hz version (with a spatial reference grating) of these motion experiments (with a temporal reference).

2) Klein<sup>31</sup> relates the Nakayama & Silverman<sup>32</sup> single displacement experiment to a sudden contrast increment experiment. Nakayama & Silverman claim that the displacement detection and discrimination data can be explained by a motion mechanism that saturates at low contrast. If Nakayama & Silverman<sup>32</sup> are correct then the Test-Pedestal approach would be confronted with a counterexample. Klein<sup>31</sup> argues that this is not the case. Rather, I claim that the motion system has the same dependence on contrast as the contrast discrimination mechanism. The experiments reported by Beard, et al.<sup>30</sup> provide confirmatory evidence for this point of view.

3) An experiment by Stromeyer, et al.<sup>33</sup> similar to that of Beard, et al.<sup>30</sup> does at first seem to provide strong evidence against the Test-Pedestal approach. Stromeyer, et al. used a counterphase test pattern, similar to the Beard, et al. experiment. However, instead of a stationary pedestal they used a counterphase pedestal of the same spatio-temporal frequency as the test,  $\cos(fx)\cos(\omega t)$ . They found that when the test was in-phase with the pedestal, thresholds exhibited standard Weber's law masking similar to Legge & Foley<sup>22</sup>. However, when the test grating had a 90 deg phase shift both in space and in time,  $\sin(fx)\sin(\omega t)$ , then the test pattern was facilitated by the pedestal even at high pedestal contrasts. For "high velocity" counterphase gratings of low spatial frequency (.5 c/deg) and high temporal frequency (20 Hz) the in-phase thresholds were more than 10 times larger than the out-of-phase thresholds. This is dramatically different from the equal thresholds found by Beard, et al. when the pedestal had zero Hz.

Why is the Stromeyer et al. experiment<sup>33</sup> different from all the others we have been considering? The answer can be most easily seen by looking at the Fourier structure of the stimulus. The pedestal consists of a rightward plus a leftward moving grating. The in-phase test consists of an increment in the contrast of both rightward and leftward components. The out-of-phase test consists of an increment to the rightward grating and a decrement to the leftward grating. The observer sees this test pattern as rapid motion to the right. An opponent motion mechanism would be blind to the pedestal and would have a facilitated response to the out-of-phase test (the incremented rightward and decremented leftward components would summate). In none of the other stimuli examined in this paper has there been such a clear difference in Fourier composition between the two stimuli being compared.

It is important to point out that the out-of-phase thresholds for the "high velocity" stimulus can be well predicted by assuming that there is no masking. The facilitated thresholds are very close to the thresholds that are found at the bottom of the dipper function for the in-phase stimulus. The Test-Pedestal approach is thus able to account for the thresholds once one gains an understanding of when does the pedestal mask the test.

The Stromeier, et al. experiment<sup>33</sup> reminds us of the original theme of this paper: the need for improved models of masking.

#### 4. CONCLUSION

This article began by discussing the need for greater involvement of vision researchers in the development of a fidelity metric. This metric would be used to assess the visible degradation of images and image sequences after they have been compressed and decompressed. The importance of developing a high quality fidelity metric can not be overstated since future visual information will be digital and will require compression.

A fidelity metric is a beautiful example of the Test-Pedestal approach. The task for the observer (and the fidelity metric) is to detect a test pattern (the difference between the displayed image and the intended image) in the presence of a pedestal (the displayed image). The Test-Pedestal approach breaks the task up into two parts: 1) the detection of the test pattern on a uniform field, and 2) the amount of masking by the pedestal. There are still important improvements to be made in both parts. As discussed in this paper the Test-Pedestal approach has succeeded in a number of domains to provide a framework for predicting discrimination thresholds of suprathreshold stimuli. We also pointed out the need to treat this approach not as an end, but rather as a tool to be used for the goal of improving filter models of vision.

#### 5. ACKNOWLEDGMENTS

This research was supported by grant R01 EY04776 from the National Eye Institute. I thank Charles Stromeier, Dennis Levi, Thom Carney, Qingmin Hu and Tina Beard for their central contributions to this research.

#### 6. REFERENCES

- <sup>1</sup> Descartes, R. Treatise of Man. Cambridge MA: Harvard Univ. Press. 1664, reprinted in 1972.
- <sup>2</sup> Marr, D. Vision. San Francisco: W.H. Freeman & Co. 1982.
- <sup>3</sup>Klein, S. A., Casson, E. & Carney, T. "Vernier acuity as line and dipole detection," *Vision Res.*, vol. 30, pp 1703-1719, 1990.
- <sup>4</sup>Klein, S. A. & Levi, D. M. "Hyperacuity thresholds of one second: Theoretical predictions and empirical validation," *J. Opt. Soc. Amer. A* vol. 2, pp 1170-1190, 1985.
- <sup>5</sup> Wilson, H. R. "Responses of spatial mechanisms can explain hyperacuity," *Vision Res.*, vol. 26, pp 453-469, 1986.
- <sup>6</sup> Klein, S. A. "Spatial vision models: Problems and successes," In Spatial Vision in Humans and Robots (pp. 3-26). Cambridge University Press, 1993.
- <sup>7</sup>Geisler, W. S. & Davila, K. D. "Ideal discriminators in spatial vision: two-point stimuli," *J. Opt. Soc. Amer. A*, vol. 2, pp 1483 - 1497, 1985.
- <sup>8</sup>Hu, Q., Klein, S. A. & Carney, T. "Can sinusoidal vernier acuity be predicted by contrast discrimination," *Vision Res.*, vol. 33, pp 1241-1258, 1993.
- <sup>9</sup>Carney, T., Klein, S. A. & Levi, D. "Resolution as dipole and quadrupole detection," *Opt. Soc. of Am. Technical Digest*, vol. 11, pp 80, 1988.
- <sup>10</sup>Levi, D. M. & Klein, S. A. "The role of local contrast in the visual deficits of humans with naturally occurring amblyopia," *Neuroscience Letters*, v. 136, 63-66, 1992.
- <sup>11</sup>Levi, D. M., Klein, S. A. & Wang, H. "The role of visibility and spatial pooling in limiting vernier acuity in amblyopic and peripheral vision," Submitted for publication, 1993.
- <sup>12</sup>Levi, D. M., Klein, S. A. & Wang, H. "Discrimination of position and contrast in amblyopic and peripheral vision," Submitted for publication, 1993.
- <sup>13</sup>Levi, D. M., Klein, S. A. & Aitsebaomo, P. "Vernier acuity, crowding and cortical magnification," *Vision Res.*, vol. 25, pp 963-977, 1985.
- <sup>14</sup>Levi, D. M. & Klein, S. A. "Vernier acuity, crowding and amblyopia," *Vision Res.*, v. 25, pp 979-991, 1985.
- <sup>15</sup>Levi, D. M. & Klein, S. A. "The role of separation and eccentricity in encoding position," *Vision Res.*, vol. 30, pp 557-585, 1990.
- <sup>16</sup>Yap, Y. L., Levi, D. M. & Klein, S. A. "Peripheral hyperacuity: isoeccentric bisection is better than radial bisection," *J. Opt. Soc. Am. A.*, vol. 4, pp 1562-1567, 1987.
- <sup>17</sup>Klein, S. A. & Levi, D. M. "Position sense of the peripheral retina," *J. Opt. Soc. Amer. A*, vol. 4, pp 1543-1553, 1987.
- <sup>18</sup>DeValois, R. L & DeValois, K. K. Spatial Vision. Oxford: Oxford Univ. Press, 1988.
- <sup>19</sup>Campbell, F. W. & Robson, J. G. "Application of fourier analysis to the visibility of gratings," *J. Physiol.*, vol. 197, pp 551-566, 1968.
- <sup>20</sup>Stromeier, C. F. & Klein, S. "Spatial frequency channels in human vision as asymmetric (edge) mechanisms," *Vision Res.*, v. 14, pp 1409-1420, 1974.

<sup>21</sup>Nachmias, J. & Sansbury, R. V. "Grating contrast: Discrimination may be better than detection," *Vision Res.*, vol. 14, pp 1039 - 1042, 1974.

<sup>22</sup>Legge, G. E. & Foley, J. M. "Contrast masking in human vision," *Vision Res.*, vol. 70, pp 1458-1471, 1980.

<sup>23</sup>Wilson, H. R. "Responses of spatial mechanisms can explain hyperacuity," *Vision Res.*, vol. 26, pp 453-469, 1986.

<sup>24</sup>Hu, Q., Klein, S. A., & Carney, T. "Masking of high spatial frequency information after a scene cut," In Society for Information Display 93 Digest, vol. 23 pp. 521-523, 1993.

<sup>25</sup>Nachmias, J., & Kocher, E. C. "Visual detection and discrimination of luminance increments," *J. Opt. Soc. Am.*, vol. 60, pp 382-389, 1970.

<sup>26</sup>Klein, S. A. "Double judgment psychophysics: problems and solutions," *J. Opt. Soc. Amer. A*, vol. 2, pp 1568-1585, 1985.

<sup>27</sup>Guinness Book of World Records. D. McFarlan (Ed.), pp. 27. New York: Bantam Books, 1991.

<sup>28</sup>Stromeyer, C. F. I., & Klein, S. A. "Evidence against narrow-band spatial frequency channels: Detectability of frequency modulated gratings," *Vision Res.*, vol. 15, pp 899-910, 1975.

<sup>29</sup>Klein, S. "Visual multipoles and the assessment of visual sensitivity to displayed images," In SPIE Proceedings on Human Vision, Visual Processing, and Digital Display vol. 1077, pp. 83-92, 1989.

<sup>30</sup>Beard, B. L., Klein, S. A., & Carney, T. "Motion thresholds can be predicted from contrast discrimination," Submitted to *Vision Res.* 1993.

<sup>31</sup>Klein, S. A. "Do motion mechanisms saturate," Submitted to *Vision Res.* 1993.

<sup>32</sup>Nakayama, K. & Silverman, G. H. "Detection and discrimination of sinusoidal grating displacements," *J. Opt. Soc. Am. A*, vol. 2, pp 267-274, 1985.

<sup>33</sup>Stromeyer, C. F., Kronauer, R. E., Madsen, J. C. & Klein, S. A. "Opponent-movement mechanisms in human vision," *J. Opt. Soc. Am. A*, vol. 1, pp 876-884, 1984.

# Neural processing of biological motion in the macaque temporal cortex

Mike W. Oram and David I. Perrett

University of St Andrews, School of Psychology, St. Andrews, Scotland, KY16 9JU, U.K.

## ABSTRACT

Cells have been found in the superior temporal polysensory area (STPa) of the macaque temporal cortex which are selectively responsive to the sight of particular whole body movements (e.g. walking) under normal lighting. These cells typically discriminate the direction of walking and the view of the body (e.g. left profile walking left). We investigated the extent to which these cells are responsive under 'biological motion' conditions where the form of the body is defined only by the movement of light patches attached to the points of limb articulation. One third of the cells (25/72) selective for the form and motion of walking bodies, showed sensitivity to the moving light displays. 7 of these cells showed only partial sensitivity to form from motion, in so far as the cells responded more to moving light displays than to moving controls but failed to discriminate body view. These 7 cells exhibited directional selectivity. 18 cells showed statistical discrimination for both direction of movement and body view under biological motion conditions. Most of these cells showed reduced responses to the impoverished moving light stimuli compared to full light conditions. The 18 cells were thus sensitive to detailed form information (body view) from the pattern of articulating motion. Cellular processing of the global pattern of articulation was indicated by the observations that none of the cells were found sensitive to movement of individual limbs and that jumbling the pattern of moving limbs reduced response magnitude. The cell responses thus provide direct evidence for neural mechanisms computing form from non-rigid motion. The selectivity of the cells was for body view, specific direction and specific type of body motion presented by moving light displays and is not predicted by many current computational approaches to the extraction of form from motion.

## 1. INTRODUCTION

Johansson (1973)<sup>1</sup> found that subjects had no difficulty in interpreting extremely impoverished images of human walking where only small light sources attached to the points of articulation (e.g. the shoulders, elbows and wrists) were visible. Indeed the interpretation of these biological motion or moving light display stimuli was reported as being "immediate and compelling"<sup>1</sup>. Subsequent investigations have revealed that human subjects can perceive a variety of information from such biological motion stimuli including dancing, running, walking, identity, gender and even sign language and facial expression<sup>2-12</sup>. The ability to interpret biological motion stimuli is present even in very young infants<sup>13,14</sup>.

Despite the rich source of information from such perceptual studies little is known about the underlying neuronal mechanisms. The similarity of behavioural performance between human and macaque subjects in processing form from motion<sup>15</sup> suggests that the macaque is a suitable model for investigating the underlying neural mechanisms of form from motion

processing. In this article we present a quantitative analysis of neuronal populations in the macaque monkey which might support the analysis of form from biological motion and compare current computational approaches to the implications of the neurophysiological findings.

### 1.1 Form and Motion Pathways

Processing of visual information in primates is thought to follow two pathways: the ventral "form" and the dorsal "motion" pathway<sup>16-18</sup>, although the degree of independence of these pathways is debated<sup>19</sup>. The ventral pathway passes through the areas V1, V2, V4, into inferotemporal cortex (IT) and the anterior sections of superior temporal sulcus (including area STPa) whereas the dorsal pathway flows from V1 through V2, the middle temporal area (MT), also known as V5, and the medial superior temporal areas (MSTl and MSTd) and then to the frontal eye fields and parietal cortex. The two pathways are not completely separate: outputs from areas MSTl and MSTd also pass through the fundus of the superior temporal sulcus (FST) to the posterior and anterior sections of the superior temporal polysensory area (STPp and STPa)<sup>20,21</sup>. Area STPa therefore receives

inputs from both the ventral (form) and dorsal (motion) pathways<sup>20,21</sup>. In view of this anatomical convergence, it may not be surprising that some neurons in area STPa (and more generally throughout the anterior sections of the superior temporal sulcus, STS) show selectivity both for the form and the direction of motion of objects including walking and articulation of individual limbs<sup>9,22-25</sup> and hand actions (e.g. tearing, object manipulation)<sup>26,27</sup>

### **1.2 Possible mechanisms of sensitivity to form and motion**

The observed conjoint selectivity to both form and motion information of single cells in STPa could be achieved by (a) simple integration of overall displacement (e.g. motion) with information of static form, or (b) by combining outputs from cells selective to the same body form but having sensitivity to different spatial locations or (c) by performing a "form from motion" analysis from motion inputs alone. Cell selectivity suitable for all three processing schemes is well documented. Head and body information is coded in both IT, which projects to STPa, and within STPa itself<sup>23,24,28-33</sup>. Areas STPa and STPp, which sends direct input to STPa both contain neurons sensitive to direction of motion, but not object form<sup>25,28,34-37</sup>. Therefore any of the outlined processing schemes could in principle be implemented by cells in area STPa, either using inputs from cells in preceding areas (IT and STPp) or using inputs from cells within STPa itself. Sensitivity to biological motion stimuli would only be seen under scheme (c) where the organisation of the motion inputs was such that sufficient information was available for performing the "form from motion" analysis.

### **1.3 View and direction specificity**

Studies of static form selectivity in areas STPa and IT strongly suggest that objects are processed and coded in a view specific manner. This finding is consistent with the finding of view specific coding of objects in STPa and IT<sup>23,24,28,31,33</sup>, work with cells selective for body form revealed preferential coding of four "characteristic" views: the face, the left and right profiles and the back view of the head and body<sup>32</sup>. Similarly, processing and coding of motion independent of form in STPa appears to be conducted in a direction specific manner, with the directions along the cartesian axes (towards, away, left, right, up, down) being preferentially represented<sup>37</sup>.

The selectivity of cells conjointly sensitive to body form and motion in STPa and other regions of the macaque temporal lobe (e.g. the amygdala<sup>22</sup>) also appears to be specific: some cells respond selectively to the left profile body view walking to the observer's left but not other view and direction combinations<sup>8,9,22,25</sup>. As mirror image body views are identical in size, complexity of articulating elements and angular speed of component movements, they have been used in psychophysical experiments for quantitative assessment of human perceptual sensitivity to biological motion stimuli<sup>5,8,10</sup>. The specificity of STPa cells sensitive to form and motion likewise allows meaningful comparison of cell responses to mirror image body views moving in the same direction.

### **1.4 Computational approaches for interpretation of biological motion stimuli**

The majority of the computational models analysing form from motion stimuli in general<sup>38,39</sup> and biological motion in particular (e.g. 40-44) establish a linkage structure that is consistent for the motion of the moving elements. This means that the analysis establishes a description that is independent of body view and direction of motion: indeed the analysis is applicable to any articulating entity, and gives no information about direction of movement nor the identity of the object that is moving. While analysis of overall displacement of an object is simple to perform, there would still be the problem of binding the direction of motion with the particular object. Furthermore the models that establish only linkage structure would require a further processing stage for determination of object identity and body view. Other computational approaches use a template of the object's identity<sup>45,46</sup>. However, these models also predict invariance with respect to orientation, body view and direction of motion (assumptions are often required about the nature of the motion - walking, running etc - in these models as well). Therefore to explain human observer performance further processing stages would also be required, in particular to extract body view. It would only be during this subsequent stage that effects such as increased response latency seen with inversion of stimuli<sup>6,47</sup>. View sensitivity of cells responsive to body motion defined under biological motion is therefore an important attribute to quantify since it is a property that is not predicted on the basis of most current computational approaches to biological motion.

## 2. METHODS

Four subjects were used (*Macaca mulatta*, 3 male B, D, H wt. 5.8 kg, 1 female J wt. 4 kg from a UK registered breeding colony). The subjects were trained to respond differentially depending on the colour of a LED attached to a plain white wall 4 m away. A half second warning tone was given before each trial, then the LED was turned on. The colour of the LED was varied in pseudo-random order across trials under computer control. Video-disc sequences or real 3-D moving objects were presented either to cross the LED or projected to cover the LED at each trial.

Standard chronic recording techniques were used to record from single cells in the STPa when stimuli were presented. Spikes from individual cells were discriminated using a threshold voltage window. The data were stored in 5 ms time bins. Responses were measured as spike frequency estimated from the period 100 - 350 ms post-stimulus onset. Eye movements were recorded (modified ACS infra-red reflection system) and sampled at the same frequency as the spike signals with 8 bit accuracy over the range +/- 20 degrees and stored with the spike data for each trial.

The stimuli were either real 3-D presentations or sequences of frames on a video disc. They included images of the experimenter walking, both forwards (compatible movement) and backwards (incompatible movement) in different directions (towards, away, left and right with respect to the subject) under normal lighting and biological motion conditions. The biological motion stimuli were made using luminescent patches (subtending approximately 0.2 degrees) fixed to the experimenter at the neck, shoulders, elbows, wrists, hips, knees and ankles. In addition to dot stimuli, stick figure representations were also used. These were generated in an analogous fashion to the biological motion stimuli but short strips of luminescent material were fixed between the articulation points. Control objects moving in the same directions as the walking and biological motion stimuli were used. These were of similar size to and like the walking stimuli had non-rigid motion. The controls used for the biological motion were luminous dots moving either rigidly or non-rigidly under blackout conditions and thirdly a 'jumbled' biological motion. The jumbled figure was made by randomly moving the coordinates of the digitized points of articulation using a computer based system (IRIS 3130, Silicon Graphics). These coordinates were moved by 30 % of the initial

head to floor height of the figure. The resulting linkage structure was unchanged but, when replayed, the image was no longer recognizable as a walking image.

Stimuli were viewed through either a liquid crystal shutter (Screen Print Technology) or a large aperture camera shutter (Compur, 6.5 cm diameter). Both shutters had rise times of < 15 ms. Each stimulus was presented 5 or more times in computer controlled pseudo-random order. Each trial consisted of a 0.5 s warning tone, followed by a 1 s stimulus presentation period. The inter-trial interval was randomly varied between 0.5 and 5 seconds. Cells were tested for selectivity to the single limb movements present in the preferred stimulus (e.g. leg and arm flexing) to ensure that the selectivity was for whole body motion. Normal lighting and biological motion conditions were tested with at least 2 body views and controls moving in 1 or 2 directions. A cell that failed to respond to stick figures was assumed to be unresponsive to dot figures.

A cell was classified as selective for walking if under normal lighting conditions there was a significant difference between one direction/body view combination from (i) control objects, (ii) a second body view moving in the same direction and if tested (iii) the same body view moving in a second direction. All the cells reported here were not found to be selective for single limb articulation but rather required whole body motion. Off line analysis for all cells took the form of 2-way ANOVA, with body view as one factor and stimulus type (natural, biological motion) as the second. When data was available a second 2-way ANOVA was performed with the direction of motion as one factor and the stimulus type (natural, biological motion, control) as the second factor. Significance for all statistical tests was taken at the 0.05 level. Post-hoc testing of the ANOVAs was performed using the protected least significant difference (PLSD) test<sup>48</sup>.

In order to examine the discrimination shown by the responses of the tested cells to the differing stimuli, discrimination measures were calculated for overall direction of motion ( $I_d$ ) and view ( $I_v$ ). Where

$$I_d = 1 - (R_{oppd}/R_{pref}),$$
$$\text{and} \quad I_v = 1 - (R_{oppv}/R_{pref}).$$

$R_{pref}$  = response to the preferred view and direction combination - spontaneous activity (SA),  $R_{oppd}$  = response to the preferred view moving in the opposite direction from the preferred direction - SA and  $R_{oppv}$  = the response to the view opposite to the

preferred view moving in the preferred direction - SA)

The preferred direction and view were first defined under normal lighting and then the magnitudes of the responses  $R_{pref}$ ,  $R_{oppv}$  and  $R_{oppd}$  measured and compared under biological motion.

### 3. RESULTS

#### 3.1 Cells selective to human walking

Of the 72 cells found to be selective for the walking stimuli that were tested for sensitivity to biological motion, 47 (65%) gave no response above spontaneous activity or control response levels. Thus approximately two thirds of the cells selective for walking bodies did not show any responsiveness to motion information alone. A further 7 cells (10%) showed a response pattern with moving light stimuli whereby these cells responded more strongly to both body views moving in the cell's preferred direction than to controls moving in the same direction, spontaneous activity or biological motion in the null (opposite) direction. Although sensitivity to body view was not seen in these cells, responses to biological motion stimuli moving in the preferred direction were greater than responses to controls moving in the same direction. This indicates partial sensitivity to body form. No cells were found with the converse selectivity - that is, showing view selectivity but not directional selectivity under moving light displays.

The remaining 18 cells out of the 72 tested (25%) showed selectivity for both form (body view) and direction of motion for biological motion of dot and stick figures. Most of these cells (14/18, 78%) showed a reduction in absolute response magnitude relative to the walking stimuli under natural lighting. Cells sensitive to both moving dot and stick figure stimuli were found with this type of response characteristic. Not surprisingly, cell responses showing selectivity for both view and direction showed relatively large responses compared with cells that were unresponsive to biological motion stimuli (Figure 1).

Four cells (22% of cells responding to biological motion stimuli responded to the biological motion stimuli in a manner that was very similar to the responses to the real walking stimuli. Figure 2 shows the responses of one cell to real and dot walking figures. As

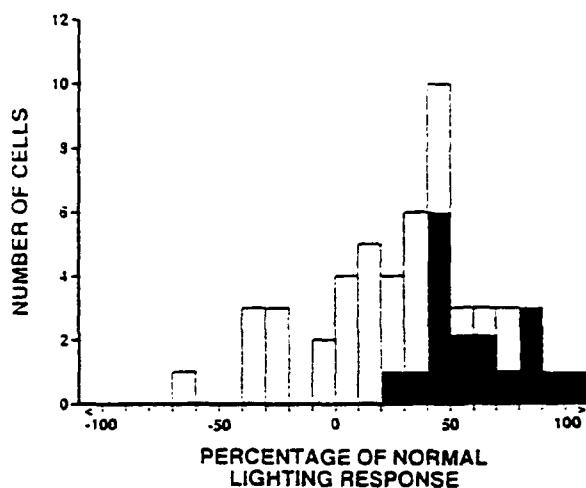


Figure 1. Responses to biological motion stimuli compared to normal lighting stimuli. Distribution of response magnitudes to the preferred view and direction combination to biological motion stimuli expressed as a percentage of the response to the same stimulus defined under normal lighting. Light bars: all cells tested. Dark bars: cells selective for view and direction under biological motion conditions.

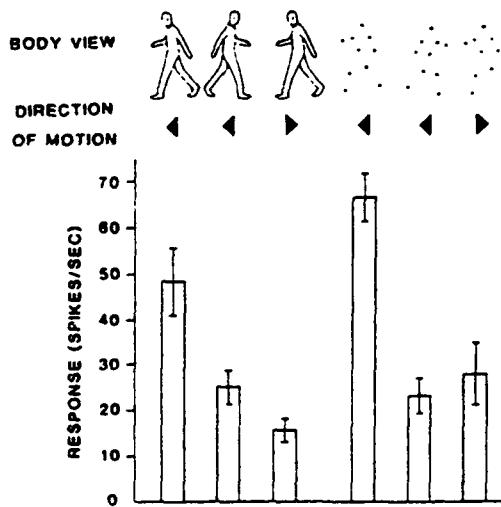


Figure 2. Responsiveness to biological motion. The mean responses +/- S.E.M. are shown to the stimuli depicted above for one cell. The cell's selectivity for compatible walking to the left is maintained with biological motion stimuli. 2-way ANOVA showed a main effect for view/direction combination ( $F_{[3,32]} = 14.0$ ,  $p < 0.0005$ ) but not lighting conditions (natural vs biological motion dots) ( $F_{[3,32]} = 3.1$ ,  $p = 0.09$ ). The interaction was non-significant ( $F_{[3,32]} = 1.16$ ,  $p = 0.34$ ). (Adapted from Oram and Perrett 1993)

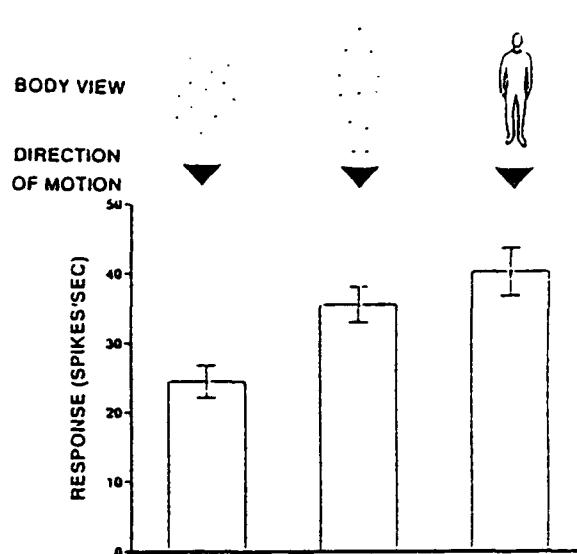
can be seen, the cell has a preferred stimulus of compatible walking to the monkey's left. The right profile walking in the preferred direction and the preferred body view walking to the right both produced significantly weaker responses. The dot figure responses also followed this pattern, with significant differences between the preferred view/direction combination and the other combinations tested under biological motion conditions.

### 3.2 Jumbled articulation

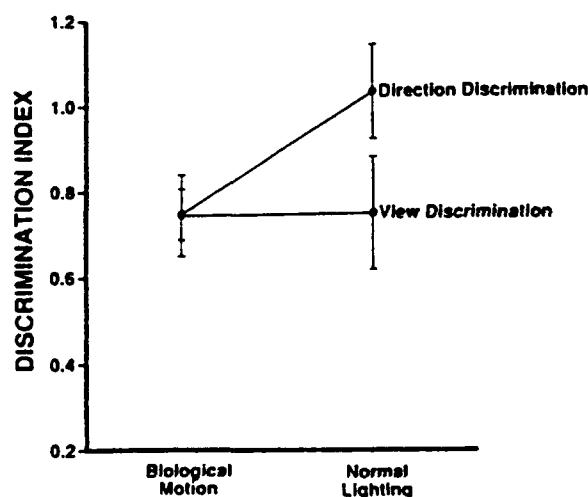
As shown in Figure 1, cell responses sensitive to biological motion showed both direction and form (body view) selectivity. As noted in the introduction, a common feature of computational models of moving light displays show no sensitivity to form or direction, but rather establish linkage structure. When combined with overall translation, the motion of individual limbs becomes approximately 180 degrees out of phase when walking in the same direction but with the opposite body view. The jumbled figure stimulus (see methods) has random "limb" movement vectors, and shows no consistent phase relationship with normal walking. Moreover, these stimuli have the same linkage structure, differing only in the relative vectors of point light movement. The jumbled figure stimuli therefore present a second control for investigation of form selectivity under biological motion conditions. Figure 3 shows the response of a cell to biological motion stimuli and a jumbled biological motion stimulus. As can be seen, the response differentiates the preferred movement (front body view walking towards) defined under either normal lighting or biological motion from the jumbled biological motion "front view" moving towards the subject.

### 3.3 Discrimination measures

A comparison of  $I_d$  and  $I_v$  values obtained under both normal and biological motion conditions is shown for those cells tested in all conditions which showed statistical discrimination for direction and view under biological motion in figure 4. A 2-way ANOVA was performed, with cell as one factor and discrimination index ( $I_d$  and  $I_v$ ) as the second. The results showed that there was a significant drop in the direction discrimination evident in the cells response when changing from normal lighting to biological motion testing conditions. Surprisingly there was no such drop seen for the view discrimination.



**Figure 3. Discrimination between normal and jumbled articulation.** The upper section depicts the stimuli. Under normal and biological motion conditions the cell responded selectively to the front body view moving towards the subject. Responses to the jumbled front view biological motion stimulus moving towards the subject was less ( $p < 0.025$ ) than the response to either non-jumbled stimulus.



**Figure 4. View and direction discrimination.** The mean (+/- SEM) of the view and direction discrimination indices of all cells selective for biological motion stimuli are plotted for normal and biological motion stimuli. While there was no significant drop between stimulus type for view discrimination, there was for direction discrimination ( $p < 0.05$ ).

## **4. DISCUSSION**

### **4.1 Summary of the results**

Of the one third of the cells (25/72) selective for the form and motion of walking bodies that were tested showed partial or full sensitivity to biological motion (either dot or stripe) stimuli, 7 cells showed selectivity for body form defined by motion independent of perspective view. These cells therefore showed a limited capacity to process 'body form' under biological motion conditions compared to full lighting conditions.

The majority (18/25) of cells responding to biological motion stimuli showed the capacity to discriminate both overall direction of movement and body view. Although these cells typically showed reduced responses to the biological motion stimuli (compared to normal lighting conditions), this reduction is perhaps not surprising given the loss of contour information. A few (4) of these cells showed response magnitudes (and selectivity) to biological motion stimuli that mimicked those to the 'real' stimuli. All 18 cells were thus sensitive to detailed form information (body view) from the pattern of articulating motion present in biological motion stimuli. Eye movement recordings under both normal and biological motion conditions were found to be comparable and could not therefore account for the differential responses<sup>64</sup>. The cell responses thus provide direct evidence for neural mechanisms computing body form from non-rigid motion.

### **4.2 Sensitivity to global motion patterns**

The cells reported here responded only to whole body motion and not simple limb articulation. It is unlikely therefore that the sensitivity observed to biological motion stimuli can be accounted for in terms of isolated patterns of local relative motion since body view was discriminated. The global nature of the selectivity was further indicated by the discrimination of whole body movements in the same direction from (1) control patterns of dots moving non-rigidly and (2) jumbled biological motion stimuli moving in the same direction. These observations indicate the complexity of the analysis being performed, since all connected pairwise relative motions of individual limbs remain in the jumbled and opposite view stimuli, yet cells did not

respond. [This type of precise selectivity also exists for many cells selective for static views of the head: these cells discriminate between different views with the same facial features (e.g. left and right profile) and they also respond less to the presentation of a jumbled face even when all the facial features are present<sup>29,32,33,49</sup>.

### **4.3 Possible motion processing in the ventral visual areas**

The involvement of the ventral stream in processing static form is indicated by the finding of single cells which exhibit a high degree of selectivity for static objects (see introduction). Indeed lesion studies have indicated that temporal cortex is needed for the learning and memory of static patterns<sup>18,50</sup>. We have shown that neural sensitivity to form and motion does not depend solely on form visible at any particular instant but can be generated from motion information alone. The computation of form from motion may well involve or depend on processing conducted in the dorsal stream of processing. Certainly lesions to the dorsal system (MT/MST) can produce impairment in the extraction of shape from motion<sup>51,52</sup>. The cell properties studied here could well depend upon the projections<sup>20</sup> from the motion processing areas (MT/MST/FST) into the cortex of the STS.

The ventral stream does however seem to be able to utilise motion processing to some extent. Lesions of the inferior temporal cortex in monkey impair the ability to learn shape discrimination where shape is defined by the relative translation of random dot patterns<sup>53</sup>. This finding suggests the processing of movement information within the ventral stream, and indeed recent studies in V2 and IT have shown that these areas are sensitive to motion defined contours and simple shapes<sup>54,55</sup>. These new findings therefore suggest that selectivity to form from motion might be achieved through the processing of contours defined by motion in the ventral stream. However, if this were the case, then cells selective for biological motion stimuli should also respond well to static images. All the cells contributing to the summary given here responded only to moving stimuli. Furthermore, testing of cells selective for static images of the head and body has not revealed selectivity to biological motion defined form (unpublished observations). It therefore seems unlikely that cell selectivity to biological motion stimuli is achieved through processing of motion defined contours along the ventral stream.

It is becoming increasingly apparent from human neuropsychological studies that recognition using static and dynamic visual cues is dissociable. Impairments in the ability to recognize facial expression from static photographs<sup>56</sup> do not necessarily parallel recognition impairments for expression displayed in biological motion format with light dots attached to the face<sup>57-59</sup>.

Neuropsychological studies also indicate that human brain mechanisms involved in the processing of complex motion (such as body form defined by biological motion and the form of cylinders defined by rigid rotation) can be dissociated from mechanisms involved in processing of direction and velocity<sup>60</sup>. More dorsal lesions are associated with a loss of simple motion processing, whereas lesions more anterior and ventral are associated with disruption of form from motion. A further example of this dissociation is provided by Patient LM<sup>61</sup> who has been described as 'motion blind' following lesions to dorsal visual areas. LM cannot track movements at velocities greater than 8 degrees per second; fast moving objects appear to her as a series of static images. Despite this dramatic motion processing deficit LM retains some capacity to recognize body form defined by biological motion stimuli (McLeod, Zihl, Perrett and Benson, unpublished studies, 1990).

#### **4.5 Relationship to computational models**

As has been pointed out by various authors, computational approaches for extraction of 3-D form from rigid motion (e.g. 39) are not applicable for interpretation of biological motion stimuli. Using correlation between moving point position and velocity, Rashid<sup>41</sup> provided one of the earliest computational model to calculate an object's linkage structure from biological motion displays. This simple procedure produced reasonable solutions for simple stimuli (an idealized walking man). For complex stimuli (e.g. 2 men walking around one another) the procedure was slow and inaccurate. Many of the more recent computational approaches (e.g. 40,42,45,62) make use of natural constraints which are likely to exist in the stimuli. For example Webb & Aggarwal<sup>44</sup> assume the axis of rotation of each locally rigid element remains fixed during the rotation. Biological motion stimuli including the walking body can be treated as a series of connected rods (one rod for the torso and two for each arm and leg). The resolved trajectory for one rod element (e.g. the torso) can be used as a frame of reference for defining the trajectory of the next linked rod element (upper arms and leg sections).

The process can be repeated until all distal rod element trajectories are defined (down to the hands or fingers as necessary). Such approaches can resolve the correct linkage in biological stimuli extremely efficiently, indeed performance can reach the theoretical limit of 3 successive frames providing no assumptions are broken<sup>42</sup>, although failure of the model's assumptions can lead to incorrect interpretations.

In summary, the assumptions made in the computational approaches outlined above are mostly about the types of motion allowed between the elements. There is a spectrum of models, each making different assumptions. They range from template matching of 3-D structure<sup>45,46</sup> to establishing the projected 2-D image linkage<sup>41</sup>. Although many models can theoretically calculate 3-D structure efficiently (within 3 snap shots or frames of motion) they are not robust but rather sensitive to failure in the assumptions (for example see Webb & Aggarwal<sup>44</sup>). Further, correspondence of light points between frames and velocity information is often assumed as part of the input data, information that is rarely available from just 3 frames<sup>41</sup>.

#### **4.5.1 The importance of view and direction specificity**

As noted in the introduction, many computational approaches to biological motion stimulus interpretation give no information about the object's identity, its direction of motion, or the perspective view. This lies in sharp contrast with the evidence from the neurophysiological data: cells respond to biological motion stimuli in a view-point specific, direction specific and object specific way. This suggests that models should calculate interpretations using view and direction specific information<sup>63</sup> to resemble more closely the biological processing of moving light displays.

Other cells found in area STPa sensitive to motion of body parts (e.g. individual limb articulation) and to different types of whole body motion (e.g. rotation and crouching) have been documented<sup>8,64</sup>. In both cases, sensitivity to view-point and direction was found. The existence of these cells suggests that computational approaches to biological motion interpretation should perform first small local view-point dependent analyses of the display (maintaining direction information). The results of these analyses should in turn be passed to a view-point, motion type and direction specific level of processing.

#### 4.5.2 Efficiency of processing

Naive observers can correctly identify a biological motion stimulus with exposure durations of between 0.1 and 0.2 seconds (4-8 frames)<sup>65,66</sup>. Subject performance on such biological motion tasks is markedly affected by the presence and type movement of background masking dots<sup>5,8,67</sup>. With computer animated biological motion displays subjects can discriminate normal walking figures from jumbled figure equivalents<sup>8</sup>. Interestingly, naive subjects perform the normal/jumble discrimination task initially rather poorly and often require more than 8 frames to perceive the figures. After minimal practice (30 trials) however, subjects can perform above chance with only 2 to 3 frames exposure, even in the presence of background masking dots, which remove residual static form cues. These human perceptual studies indicate that purely dynamic cues can be used to retrieve structure extremely quickly. Considering macaque STPa cell response latencies similar conclusions can be reached. Although detailed studies of the response time course have yet to be made, it is apparent that cell responses to biological motion stimuli can occur within 150 ms (3-4 video frames) after stimulus onset<sup>64</sup>.

#### 4.5.3 'Top-down' and 'feed-forward' influences

The improvement of human perceptual performance with practice indicates that the processing of biological motion stimuli may in some way involve 'top-down' influences where expectations for the form of the moving object are compared against visual input. The appropriate computational model for processing would appear to be one in which input data are checked against specific models stored in memory and the results of the matching used to guide subsequent predictions<sup>45,46</sup>. A role for top-down influences has also been suggested for object recognition<sup>68,69-71</sup>. It remains to be determined what role experience has in shaping STPa cell responses to biological motion stimuli.

A second problem concerning the responses of STPa cells is the source of their feed-forward motion input information. While there is a strong trend for motion processing in the dorsal stream to become more global (i.e. sensitive to the overall direction of motion), the lateral area of MST (MSTl) contains a large number of direction sensitive cells with relatively small receptive fields. The selectivity of STPa cells for the configuration of biological motion stimuli indicates that their inputs

must contain relatively local analysis of motion (e.g. limb configuration information - see above). It is therefore unlikely that global motion inputs underlie STPa cell sensitivity to biological motion.

#### 4.5.4 As associative model for biological motion sensitivity

Finally, we speculate here on a simple associative mechanism that might explain the emergence of sensitivity of some STPa cells to biological motion stimuli. While we have focused here on responses to biological motion stimuli, most cells in area STPa sensitive to walking bodies also show sensitivity to translation (in the appropriate direction) of a non-articulating image of the body (in the appropriate view)<sup>8</sup>. This sensitivity to translation is also found for many of the cells sensitive to biological motion. This suggests that cells in the STPa sensitive to walking receive form information and motion information separately. Further, the sensitivity of some biological motion sensitive cells to translating stimuli also indicates that both local and global motion inputs may influence cell activity.

Cells in STPa could receive inputs about body form (from the ventral stream) and various local motion inputs about limb articulations (from the dorsal stream). When the body locomotes, the sight of *one* body view translating in *one* direction would be associated with *one* set of local motions. This triple association would be the basis for learning the collection of articulations which typify one body view moving in one direction, with one type of action (walking).

The responses of cells to the static form of the body cluster around particular 'characteristic' views (front, back and profile)<sup>32</sup>. Further the direction selective neurons in area STPa cluster around 'characteristic' directions (left, right, towards and away)<sup>37</sup>. If sensitivity to biological motion is learned associatively then it should occur with the same type of view and direction selectivity. This is exactly what we found. The data from our studies fit several predictions from the associative learning scheme proposed here: (1) sensitivity to biological motion should show form and direction selectivity, (2) since form inputs are required as a basis for associative learning, responses to biological motion stimuli are likely to be reduced compared with normal lighting stimuli, (3) sensitivity to biological motion should not be present for cells responsive to the form of static bodies (our current data supports this - unpublished observations), (4) cells showing selectivity walking

movements under biological motion conditions should also show selectivity for translation (in the appropriate direction) of the rigid body form (in the appropriate views). The physiological evidence is thus consistent with a scheme of processing in which the local motion of limbs is associatively learned with paired form and global motion information.

### 5. ACKNOWLEDGEMENTS

This research was funded by project grants from the M.R.C., S.E.R.C., United States O.N.R. and a Royal Society University Research Fellowship to D.P.. We acknowledge the contribution of M.H. Harries, W. Dittrich, J.K. Hietanen, P.J. Benson, A.J. Chitty, A.J. Mistlin, A.S Head, who participated in some of the experiments and to P.J. Benson for the programs allowing the jumbling of the biological motion stimuli.

### 6. REFERENCES

1. G. Johansson "Visual perception of biological motion and a model for its analysis," *Percept. Psychophys.*, Vol. 14, pp. 201-211, 1973.
2. J.E. Cutting and L.T. Kozlowski "Recognising friends by their walk: Gait perception without familiarity cues," *Bulletin of the Psychon. Soc.*, Vol. 9, pp. 353-356, 1977.
3. J.E. Cutting "Generation of synthetic male and female walkers through manipulation of a biomechanical invariant," *Perception*, Vol. 7, pp. 393-405, 1978.
4. J.E. Cutting, D.R. Proffitt and L.T. Kozlowski "A biomechanical invariant for gait perception," *J. Exp. Psychol.*, Vol. 4, pp. 357-372, 1978.
5. J.E. Cutting, C. Moore and R. Morrison "Masking the motion of human gait," *Percept. Psychophys.*, Vol. 44, pp. 339-347, 1988.
6. W.D. Dittrich "Action categories and the perception of biological motion," *Perception*, Vol. 22, pp. 15-22, 1993.
7. L.T. Kozlowski and J.E. Cutting "Recognizing the sex of a walker from a dynamic point-light display," *Percept. Psychophys.*, Vol. 21, pp. 575-580, 1977.
8. D.I. Perrett, M.H. Harries, P.J. Benson, A.J. Chitty, and A.J. Mistlin "Retrieval of structure from rigid and biological motion: an analysis of the visual response of neurons in the macaque temporal cortex," In *AI and the Eye*. A Blake and T Troscianko (eds), pp 181-201, John Wiley, Chichester, England, 1990a.
9. D.I. Perrett, M.H. Harries, A.J. Chitty, and A.J. Mistlin "Three stages in the classification of body movements by visual neurons." In Images and Understanding, HB Barlow, C Blakemore and M Weston-Smith, (eds), pp 94-107, Cambridge University Press, Cambridge, England, 1990b.
10. G. Mather, K. Radford and S. West "Low level visual processing of biological motion," *Proc. R. Soc. Lond. B*, Vol. 249, pp. 149-155, 1992.
11. G. Jansson and G. Johansson "Visual perception of bending motion," *Perception*, Vol. 2, pp. 321-326, 1973.
12. H. Poizner, U. Bellugi and V. Lutes-Driscoll "Perception of American Sign Language in dynamic point-light displays," *J. Exp. Psychol. (Human Perception and Performance)*, Vol. 7, pp. 430-440, 1981.
13. B.I. Bertenthal, D.R. Proffitt, N.B. Spenter and M.A. Thomas "The development of infant sensitivity to biomechanical motions," *Child Devel.*, Vol. 56, pp. 531-543, 1985.
14. R. Fox and C. McDaniel "The perception of biological motion by human infants," *Science*, 218, pp. 486-487, 1982.
15. R.M. Siegel and R.A. Andersen "The perception of structure from visual motion in monkey and man," *J. of Cogn. Neurosci.*, Vol. 2, pp. 306-319, 1990.
16. F.A. De Yoe and D.C. Van Essen "Concurrent processing streams in monkey visual cortex," *TINS*, Vol. 11, pp. 219-226, 1988.
17. M. Mishkin, L.G. Ungerleider and K.A. Macko "Object vision and spatial vision: Two cortical pathways," *TINS*, Vol. 6, pp. 414-417, 1983.
18. L.G. Ungerleider and M. Mishkin "Two cortical visual systems," In: Analysis of visual behaviour, D.J. Ingle, M.A. Goodale and R.J.W. Mansfield (eds) MIT Press, Cambridge, MA, U.S.A., pp 549-586, 1982.
19. W.H. Merigan and J.H.R. Maunsell "How parallel are the primate visual pathways?," *Annu. Rev. Neurisci.* Vol. 16, pp 369-402, 1993.
20. D.J. Felleman and D.C. Van Essen "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex*, Vol. 1, pp. 1-47, 1991.
21. M.P. Young "Objective analysis of the topological organization of the primate cortical visual system," *Nature Lond.*, Vol. 358, pp. 152-155, 1992.
22. L. Brothers and B. King "A neuroethological framework for the representation of minds," *J. Cogn. Neurosci.*, Vol. 4, pp. 107-118, 1992.
23. R. Desimone, T.D. Albright, C.G. Gross and C. Bruce "Stimulus-selective properties of inferior temporal neurons in the macaque," *J. Neurosci.*, Vol. 8, pp. 2051-2062, 1984.

24. M.E. Hasselmo, E.T. Rolls, G.C. Baylis, and V. Nalwa "Object-centered encoding by face-selective neurons in the cortex in the superior temporal sulcus of the monkey," *Exp. Brain Res.*, Vol. 75, pp. 417-429, 1989.

25. D.I. Perrett, P.A.J. Smith, A.J. Mistlin, A.J. Chitty, A.S. Head, D.D. Potter, R. Broennimann, A.D. Milner, and M.A. Jeeves "Visual analysis of body movements by neurons in the temporal cortex of the macaque monkey: a preliminary report," *Beh. Brain Res.*, Vol. 16, pp. 153-170, 1985b.

26. D.I. Perrett, A.J. Mistlin, M.H. Harries and A.J. Chitty "Understanding the visual appearance and consequence of hand actions." In Vision and Action: The Control of Grasping. Ablex Publishing, Norwood, New Jersey, USA. pp. 163-180, 1989a.

27. D.I. Perrett, M.H. Harries, R. Bevan, S. Thomas, P.J. Benson, A.J. Mistlin, A.J. Chitty, J.K. Hietanen and J.E. Ortega "Frameworks of analysis for the neural representation of animate objects and actions," *J. Exp. Biol.*, Vol. 146, pp. 87-114, 1989b.

28. C.J. Bruce, R. Desimone and C.G. Gross "Visual properties on neurons in a polysensory area in superior temporal sulcus of the macaque," *J. Neurophysiol.*, Vol. 46, pp. 369-384, 1981.

29. D.I. Perrett, E.T. Rolls and W. Caan "Visual neurons responsive to faces in the monkey temporal cortex," *Exp. Brain Res.*, Vol. 47, pp. 329-342, 1982.

30. D.I. Perrett, P.A.J. Smith, D.D. Potter, A.J. Mistlin, A.S. Head, A.D. Milner, and M.A. Jeeves "Neurons responsive to faces in the temporal cortex: studies of functional organization sensitivity to identity and relation to perception," *Hum. Neurobiol.*, Vol. 3, pp. 197-208, 1984.

31. D.I. Perrett, P.A.J. Smith, D.D. Potter, A.J. Mistlin, A.S. Head, A.D. Milner, and M.A. Jeeves "Visual cells in the temporal cortex sensitive to face view and gaze direction," *Proc. R. Soc. Lond. B*, Vol. 223, pp. 293-317, 1985a.

32. D.I. Perrett, M.W. Oram, M.H. Harries, R. Bevan, J.K. Hietanen, P.J. Benson and S. Thomas "Viewer-centred and object-centred coding of heads in the macaque temporal cortex," *Exp. Brain Res.*, Vol. 86, pp. 159-173, 1991.

33. D.I. Perrett, J.K. Hietanen, M.W. Oram, and P.J. Benson "Organization and functions of cells responsive to faces in the temporal cortex," *Phil. Trans. R. Soc. Lond. B*, Vol. 335, pp. 23-30, 1992.

34. C.G. Gross, C.E. Rocha-Miranda and D.B. Bender "Visual properties of neurons in inferotemporal cortex of the monkey," *J. Neurophysiol.*, Vol. 35, pp. 96-111, 1972.

35. J.K. Hietanen and D.I. Perrett "Motion sensitive cells in the macaque superior temporal polysensory area: I. lack of response to the sight of the monkey's own hand," *Exp. Brain Res.*, Vol. 93, pp. 117-128, 1993.

36. K. Hikosaka, E. Iwai, H-A Saito and K. Tanaka "Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey," *J. Neurophysiol.*, Vol. 60, pp. 1615-1637, 1988.

37. M.W. Oram, D.I. Perrett, and J.K. Hietanen "Directional tuning of motion sensitive cells in the anterior superior temporal polysensory area (STPa) of the macaque," *Exp. Brain Res.*, in press.

38. E.C. Hildreth and C. Koch "The analysis of visual motion: From computational theory to neuronal mechanisms," *Ann. Rev. Neurosci.*, Vol. 10, pp. 477-533, 1987.

39. S. Ullman "The interpretation of structure from motion," *Proc. R. Soc. Lond.: Series B*, Vol. 203, pp. 405-426, 1979.

40. D.D. Hoffman and B.E. Flinchbaugh "The interpretation of biological motion," *Biol. Cybern.*, Vol. 42, pp. 195-204, 1982.

41. R.F. Rashid "Towards a system for the interpretation of moving light displays," *IEEE, PAMI*, Vol. 2, pp. 574-581, 1980.

42. N. Sugie and K. Kato "A computational model of biological motion perception," *IEEE Montech. Conference on Biomedical Technologies*, pp. 140-143, Nov. 1987.

43. K. Sugihara and N. Sugie "Recovery of rigid structure from orthographically projected optical-flow," *Computer Vision, Graphics and Image Processing*, Vol. 27, pp. 309-320, 1984.

44. J. Webb and J. Aggarwal "Structure from motion of rigid and jointed objects," *Art. Intel.*, Vol. 19, pp. 107-130, 1992.

45. H.J. Lee and Z. Chen "Determination of 3D human body postures from a single view," *Computer Vision, Graphics and Image Processing*, Vol. 30, pp. 148-168, 1985.

46. M.K. Leung and Y.H. Yang "A region-based approach for human body motion analysis," *Pattern Recognition*, Vol. 20, pp. 321-339, 1987.

47. S. Sumi "Upside-down presentation of the Johansson moving light-spot pattern," *Perception*, Vol. 13, pp. 283-286, 1984.

48. G.W. Snedecor and W.G. Cochran *Statistical Methods*. Iowa State University Press, Ames, Iowa, 1980.

49. D.I. Perrett, A.J. Mistlin, A.J. Chitty, P.A.J. Smith, D.D. Potter, R. Broennimann, and M.H. Harries "Specialized face processing and hemispheric

asymmetry in man and monkey: Evidence from single unit and reaction time studies," *Beh. Brain Res.*, Vol. 29, pp. 245-258, 1988.

50. P. Dean "Effects of inferotemporal lesions on the behaviour of monkeys," *Psychological Bulletin*, Vol. 83, pp. 41-71, 1976.
51. R.A. Andersen and R.M. Siegel "Motion processing in the primate cortex," In Signal and Sense: Local and Global Order in Perceptual Maps, G.M. Edelman, W.E. Gall & W.M. Cowan (Eds) pp. 163-184, 1989.
52. R.M. Siegel and R.A. Andersen "Motion perceptual deficits following ibotenic acid lesions of the middle temporal area in the behaving rhesus monkey," *Soc. Neurosci. Abstr.*, Vol. 12, p. 1183, 1986.
53. K.H. Britten, W.T. Newsome and R.C. Saunders "Effects of inferotemporal cortex lesions on form-from-motion discrimination in monkeys," *Exp. Brain Res.*, Vol. 88, pp. 292-302, 1992.
54. E. Peterhans and R. von der Heydt "Functional organization of area V2 in the alert macaque," *Europ. J. Neurosci.*, Vol. 5, pp. 509-524, 1993.
55. G. Sary, R. Vogels and G. Orban "Cue-invariant shape selectivity of macaque inferior temporal neurons," *Science*, Vol. 260, pp. 995-997, 1993.
56. P. Ekman and W.V. Friesen *Pictures of facial affect*. Consulting Psychologists Press, Palo Alto, U.S.A., 1976.
57. J.N. Bassili "Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face," *Journal of Personality and Social Psychology*, Vol. 37, pp. 2049-2058, 1979.
58. R. Campbell "The neuropsychology of lip-reading," *Phil. Trans. R. Soc. Lond. B*, Vol. 335, pp. 39-45, 1992.
59. G.W. Humphreys, N. Donnelly and M.J. Riddoch "Expression is computed separately from facial identity and it is computed separately for moving and static faces: Neuropsychological evidence," *Neuropsychologia*, Vol. 31, pp. 173-181, 1993.
60. L.M. Vaina, M. Lemay, D.C. Biensang, A.Y. Choi and K. Nakayama "Intact 'biological motion' and 'structure from motion' perception in a patient with impaired motion mechanisms: A case study," *Visual Neurosci.*, Vol. 5, pp. 353-369, 1990.
61. J.D. Zihl, D. Von Cramon, and N. Mai "Selective disturbance of vision after bilateral brain damage," *Brain*, Vol. 106, pp. 313-340, 1983.
62. B.M. Bennett and D.D. Hoffman "The computation of structure from fixed-axis motion: Nonrigid structures," *Biol. Cybern.*, Vol. 51, pp. 293-300, 1985.
63. N.H. Goddard "The perception of articulated motion: Recognizing moving light displays," PhD thesis, University of Rochester, New York, U.S.A., 1992.
64. M.W. Oram and D.I. Perrett "Responses of anterior superior temporal polysensory (STPa) neurons to 'biological motion' stimuli.", in submission, *J. Cogn. Neurosci.*
65. G. Johansson "Spatio-temporal differentiation and integration in visual motion perception," *Psychological Research*, Vol. 38, 379-393, 1976.
66. J.S. Lappin, J.F. Doner and B. Kottas "Minimal conditions for the visual detection of structure and motion in three dimensions," *Science*, Vol. 209, pp. 717-719, 1980.
67. D.R. Proffitt, B.I. Bertenthal and R.J. Roberts Jr "The role of occlusion in reducing multistability in moving point-light displays," *Percept. Psychophys.*, Vol. 36, pp. 315-323, 1984.
68. D.G. Lowe *Perceptual organization and visual recognition*. Kluwer Academic Publishers, Boston U.S.A., 1987.
69. M. Seibert and A.M. Waxman "Learning aspect graph representations from view sequences". In Advances in Neural Network Information Processing Systems D S Touretzky (Ed), Vol. 2, pp. 258-265, Morgan Kaufman, San Mateo, California, USA, 1991.
70. M. Seibert and A.M. Waxman "Adaptive 3D object recognition from multiple views.", *IEEE-PAMI*, Vol. 14, pp. 107-124, 1992.
71. M. Seibert and A.M. Waxman "Learning and recognizing 3D objects from multiple views in a neural system," In: Neural Networks for Perception, H. Wechsler (Ed), pp. 426-444, 1992.

# A neuronal mechanism for signaling the direction of self-motion

Jean-Pierre Roy

Montreal Neurological Institute, McGill University  
Department of Neurology and Neurosurgery  
Montreal, Quebec  
H3A 2B4

## ABSTRACT

Movement of an observer through the environment generates motion on the retina. This optic flow contains information about the direction of self-motion. To accurately signal the direction of self-motion however, the optic flow has to carry some depth information: there has to be differential motion of elements at different depths. One depth cue that is available to an organism with frontal eyes is binocular disparity. Cells in the dorsal subdivision of the Medial Superior Temporal area (area MSTd) have been proposed to play a role in the analysis of optic flow. We have examined the disparity sensitivity of neurons from MSTd in awake behaving monkeys in an attempt to understand the possible contribution of disparity to the computation of the direction of self-motion. Cells with a response to fronto-parallel motion were examined.

While the monkey looked at a fixation spot on a screen in front of it, random dot stimuli moved in the preferred direction of the cell under study, and the disparity of the dots made the stimuli appear to move in a fronto-parallel plane in front of, on, or behind the screen. Over 90% of the neurons studied were sensitive to the disparity of the visual stimulus. Of those disparity sensitive cells, 95% responded best either to near stimuli (stimuli with crossed disparities appearing to move in front of the screen) or to far stimuli (stimuli with uncrossed disparities appearing to move behind the screen).

In 40% of the disparity sensitive cells, we found cells whose preferred disparity reversed as the direction of stimulus motion was reversed. For example, a cell that responded best to crossed disparities (foreground) for rightward motion, responded best to uncrossed disparities (background) for leftward motion. Such an opposite motion of foreground and background occurs when an organism tracks a stationary object while translating in a direction different from the line of gaze.

We propose that the reversal of disparity selectivity with a reversal in direction selectivity indicates one way in which these neurons could signal the direction of self-motion of the organism in its environment.

## 1. INTRODUCTION

One function of vision is to inform the organism about its direction of heading: where it is going. It has been proposed that optic flow could provide information about the direction of self-motion.<sup>1</sup> Human studies have confirmed that, under certain circumstances, optic flow alone was sufficient for the subjects to accurately determine their direction of heading, the translational component of the self-motion.<sup>2,3,4</sup> An interesting finding of these studies is that, to accurately convey the information about the direction of heading, the optic flow stimulus has to contain depth information: there has to be differential motion of elements located at different depths. This need for differential motion of elements at different depths had been proposed earlier on theoretical grounds.<sup>5</sup>

The dorsal part of the medial superior temporal area (area MSTd) has been postulated to analyze optic flow.<sup>6,7,8</sup> Because of this postulated role in optic flow analysis, because optic flow could provide information about the direction of heading, and because depth appears crucial to the ability of accurately determining the direction of heading, we examined the response of neurons in MSTd to depth stimuli.

## 2. DISPARITY SENSITIVITY OF MSTd NEURONS

One source of information about depth in an organism with frontal eyes is binocular horizontal disparity.<sup>9</sup> We decided then to first examine the response of MSTd neurons to binocular disparity. Two hundred and seventy-two neurons from three hemispheres of two Rhesus monkeys were recorded. The neurons studied all responded in a direction selective manner to fronto-parallel motion presented on a screen where the animal fixated. The experimental conditions are as described in Roy et al.<sup>11</sup>

A cell was considered to be sensitive to disparity if it responded differently to different horizontal disparities. Figure 1 shows the response of a disparity sensitive neuron to a random dot stimulus moving in the preferred direction and at the preferred speed for this neuron.

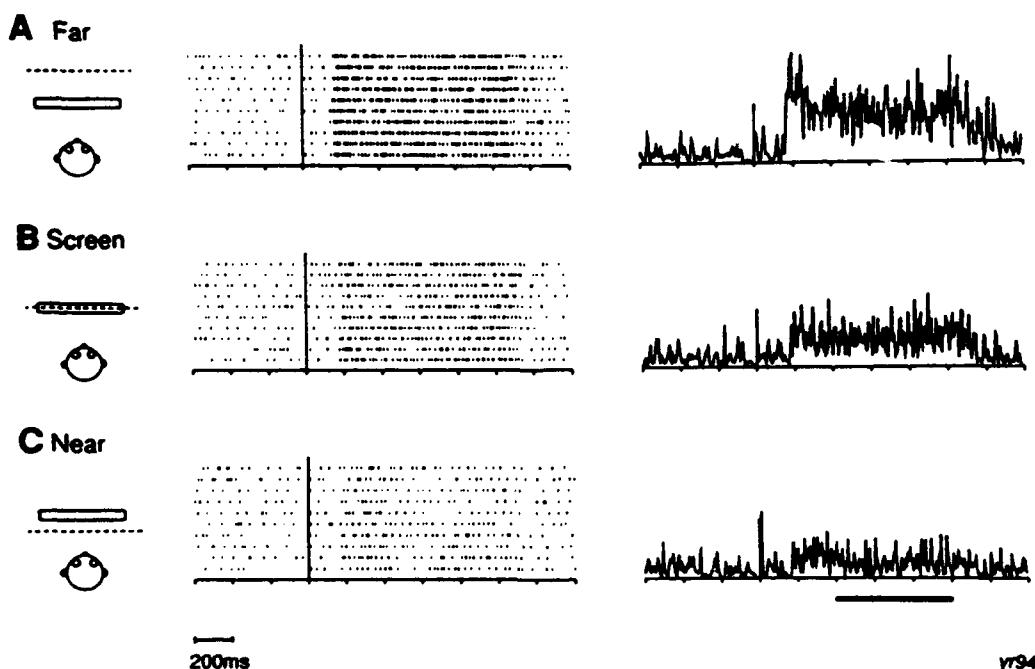


FIG. 1 Example of an MSTd neuron responding best to uncrossed disparities, Rasters (middle column) and spike density functions (right column) are shown for one cell at three different disparities. Disparity of the random dot stimulus was uncrossed  $2^\circ$  (A),  $0^\circ$  (B), or crossed  $2^\circ$  (C). The cell discharged at a high level for uncrossed disparity corresponding to motion behind the screen (Far), it responded moderately for zero disparity corresponding to motion on the screen, and it responded poorly for crossed disparity corresponding to motion in front of the screen (Near). The solid bar under the spike density function indicates the time period (400-1000 msec) after the stimulus onset (indicated by the vertical bar) over which cell discharge was counted (from Roy et al.<sup>11</sup>)

In Fig. 1A, the disparity of the moving stimulus is uncrossed  $2^\circ$ ; this corresponds to stimulus motion behind the point of fixation. The response to this disparity is an increase in discharge rate above the spontaneous level as indicated on the adjacent raster and spike density plots. The same cell gives a moderate response to a stimulus with no disparity, which corresponds to motion on the screen (Fig. 1B), and only a weak response to a stimulus of  $2^\circ$  crossed disparity, which corresponds to motion in front of the fixation point (Fig. 1C). This cell then responds best to uncrossed stimuli.

Of the disparity sensitive neurons studied in MSTd, most responded best either to crossed or to uncrossed disparities, they were either near cells or far cells using the classification of Poggio and Fischer.<sup>10</sup> Figure 2 illustrates an example of the two different disparity types. The disparity tuning curves show the mean and standard error of the tonic discharge of the cells between 400 and 1000 ms after stimulus onset.

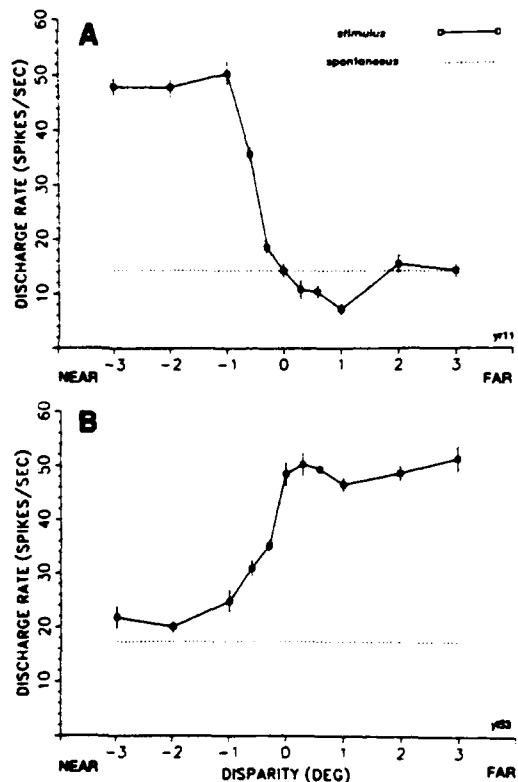


FIG. 2 Disparity sensitivity of a near cell (A) and a far cell (B). In A, the neuron responds to moving stimuli with crossed disparities. In B, the neuron responds to stimuli with uncrossed disparities. Means and SE are shown for the discharge rate at each disparity of the stimulus. The analysis of the discharge rate included the period between 400 and 1000 msec after stimulus onset. The stimuli moved in the frontoparallel plane and in the preferred direction at 11 disparities from crossed  $3^\circ$  ( $-3^\circ$ ) to uncrossed  $3^\circ$  ( $+3^\circ$ ). For both cells, each value is the mean of 10 responses. On this and all subsequent graphs, the means and SE of the spontaneous discharge rate are indicated by the dotted lines and are derived from the 600 msec period before stimulus onset (from Roy et al.<sup>11</sup>).

In Fig. 2A the disparity tuning curve for a near cell shows a discharge above the spontaneous rate for crossed disparities. The far cell shown in Fig. 2B, on the other hand, discharges above the spontaneous rate for uncrossed disparities. Tuned cells<sup>10</sup> with no near or far component were rare. A more commonly observed response was what we called a mixed response: a far or near response with a superimposed tuned response for disparities around  $0^\circ$ .

In order to determine the frequency of the types of disparity responses seen in MSTd, we classified the cells as described in Roy et al.<sup>11</sup> Over 90% of the neurons studied (228/252 cells) were sensitive to disparity. Of the disparity sensitive neurons, 95% were either near or far (216/228 cells). Of these 216 cells, 42 (19%) also had a tuned component: they were mixed cells. Pure tuned cells were rare (5%), and only 1 of the 12 tuned cells was a tuned inhibitory neuron (with cell discharge below background at 0° disparity).

### 3. DISPARITY DEPENDENT DIRECTION SELECTIVE CELLS

The next point that we examined was whether there would be a reversal of disparity selectivity for a reversal in the direction of motion of the stimuli. Consider an observer translating rightwards while tracking a stationary object in front of him. To a first approximation, images of objects in the foreground (closer than the fixation point) will move to the left, while images of objects in the background (farther than the fixation point) will move to the right (Fig. 3).

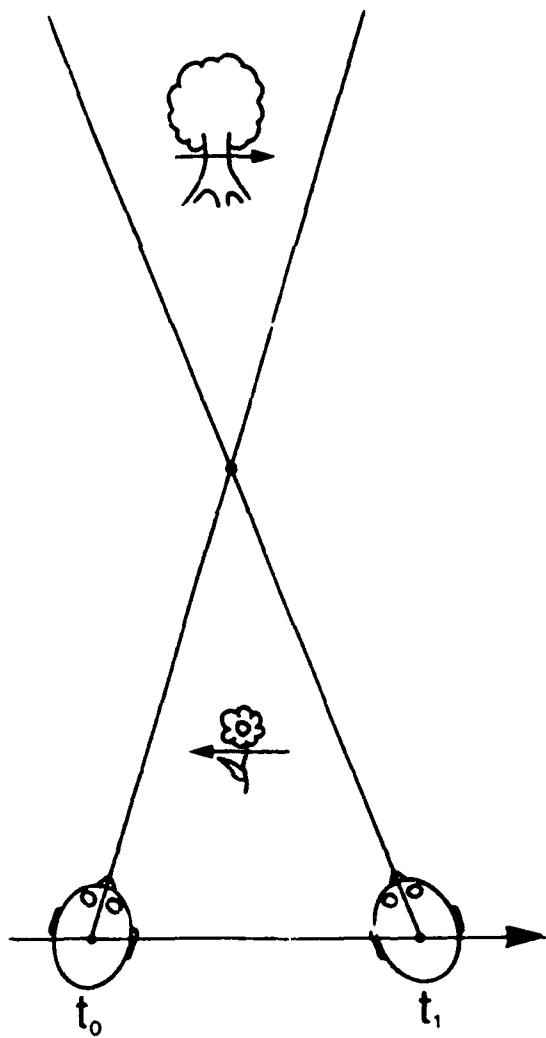


FIG. 3 Opposite motions of the background and foreground during translation (top view). As an observer translates to the right (large right-pointing arrow) while tracking an object, the direction of motion of the images of objects will depend on their depth relative to the point of fixation (the dot). An object in the background (behind the fixation point) "moves" in the direction of translation; at time  $t_0$ , the image of the tree is to the left of the line of gaze, at time  $t_1$ , it is to the right: relative to the line of gaze, its image has moved to the right (small right-pointing arrow). An object in the foreground (in front of the fixation point) "moves" in the opposite direction; at time  $t_0$ , the image of the flower is to the right of the line of gaze, at time  $t_1$ , it is to the left: its image has moved to the left (small left-pointing arrow). These opposite motions of the background and foreground will be generated when the direction of translation and the direction of tracking gaze movements are different (from Roy and Wurtz<sup>14</sup>).

Without a signal indicating the depth of the two opposite motions, the direction of translation is ambiguous: the translation of the observer could be to the left or to the right. If the direction of flow can be tagged with a depth signal however, the ambiguity about the direction of translation is removed.

To test whether the disparity signal found in MSTd neurons is appropriate to determine the direction of translation from the two opposite motions shown in Fig. 3, we presented the disparity stimuli moving in the preferred and then in the non-preferred direction for the neuron under study. Of the 65 cells studied, 39 (60%) responded best to the same sign of disparity for the two opposite directions of motion. In all these cells, one direction elicited a much stronger response to the preferred disparity than the opposite direction (Fig. 4).

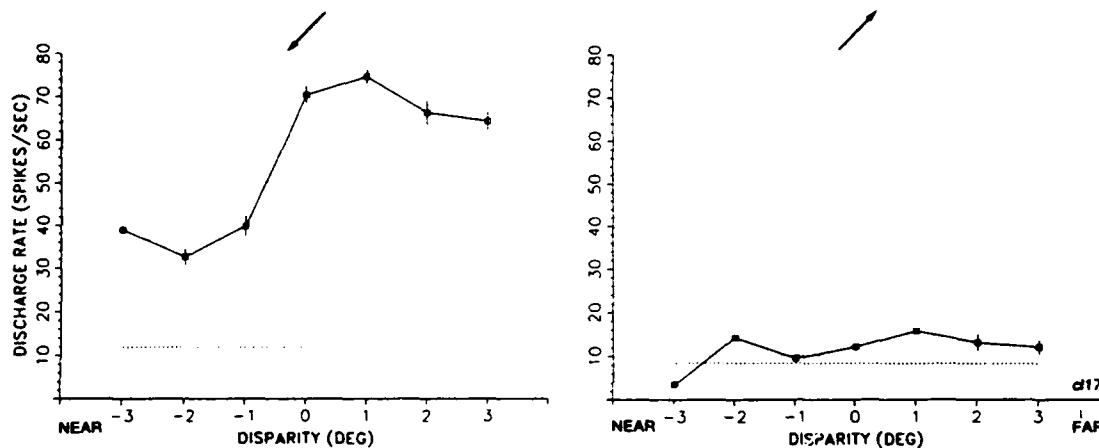
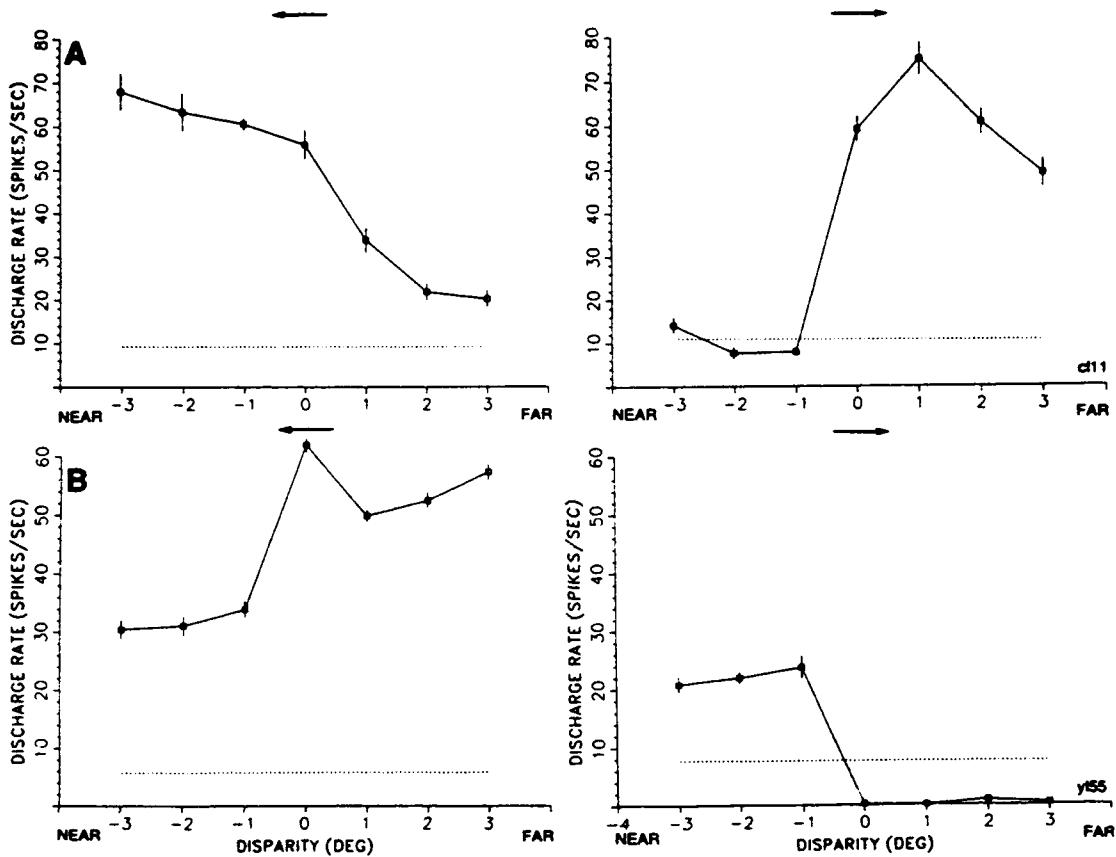


FIG. 4 Example of a non-DDD neuron. The neuron responded to crossed disparities for one direction of stimulus motion (left and down) but did not respond to motion in the opposite direction (right and up) (from Roy et al.<sup>11</sup>).

These neurons then could detect the direction of motion of the background (as in Fig. 4) or foreground. The depth signal added to the direction signal however, provides information about the direction of translation only under certain conditions. For example, if the observer looks at infinity, there will be foreground motion only, and a background responsive cell will be silent and so, would not signal the direction of translation.

We did find neurons in MSTd that appear to play a more general role in signaling the direction of translation. Twenty-six cells out of the 65 tested (40%) responded best to one direction of motion when the animal was presented with visual stimuli of one sign of disparity and the opposite direction of motion when presented with visual stimuli of the opposite sign of disparity (Fig. 5).

This disparity-dependent direction selectivity (DDD) means that the cell will respond during translation in one direction, irrespective of where in depth the animal is fixating. The neuron will respond when the foreground moves to the left or when the background moves to the right or both. These neurons then seem to signal the direction of translation relative to the object fixated under any conditions of viewing, as long as the direction of gaze and the direction of translation are different.



**FIG. 5 Examples of DDD neurons.** In *A*, the neuron responded to crossed disparities for stimuli moving to the left and the same neuron responded to uncrossed disparities for stimuli moving in the opposite direction. In *B*, the cell responded to uncrossed disparities for motion to the left but the excitatory response was to crossed disparities for motion to the right (from Roy et al.<sup>11</sup>).

If we compare the visual characteristics of DDD neurons with those of non-DDD neurons, we find no difference in receptive field size, preferred disparity, and ipsilateral or contralateral preferred directions. The only difference between DDD and non-DDD neurons was that a larger proportion of DDD cells (62%) preferred motion along the horizontal axis when compared to non-DDD cells (36%). This preference for horizontal motion in DDD neurons supports our interpretation of a role for these cells in indicating the direction of translation: macaque monkeys are primarily terrestrial animals and so their locomotion will more often be horizontal.

Until now, we have equated one given direction of motion with one given disparity as if these two parameters were the same. In fact, the two phenomena are relatively independent. The direction of motion of what is in front and what is behind the point of fixation depends exclusively on the geometry of the compensating gaze movement, a monocular phenomenon; the disparity on the other hand depends on the convergence of the two eyes on the object of fixation, a binocular phenomenon. To a first approximation, the two points, point of gaze stabilization and point of eye convergence, can be equated since under most circumstances the subject will try to track and converge on the same object. The two points are not

necessarily the same however, either because of imperfect vergence or imperfect gaze stabilization. This could account for responses at the level of the preferred disparity for 0° in some neurons. If the gaze compensation for the translation is too important for example, the object fixated will move in the direction of the translation, i.e. leave the background. Assuming perfect vergence, the object at 0° then will move in the direction of background motion. The finding that the discharge rate for 0° is more often associated with far disparities (background) would suggest that gaze overcompensation is more common than gaze undercompensation.

Another property expected of neurons that provide information about the direction of translation is that they should be sensitive to relative depth, the distance relative to the point of fixation, and not to absolute depth, the distance relative to the subject. In Fig. 3, if the subject chooses another object to fixate, both the point of vergence and the point of gaze stabilization will change together. Because of this, the disparity and direction will also vary together: leftward motion will be in the foreground, rightward motion in the background no matter what actual distances foreground/background correspond to. It appears preferable for a system signaling self-motion to carry a signal about depth relative to the point of fixation and not depth relative to the subject. Most cells in MSTd responded to pure disparity independent of the angle of vergence.<sup>11</sup>

#### 4. CONCLUSION

We propose that the DDD neurons described here have the attributes necessary for signaling the direction of translation when the observer moves in one direction while tracking an object in another direction. This corresponds to the condition where there is a rotational component to the self-motion (the tracking gaze movement) superimposed on the translational component (the locomotion). When there is no rotational component to the self-motion, such as when the observer translates directly towards the object tracked, the optic flow is a pure expansion and the DDD neurons, then, would not discharge. Other neurons that do respond specifically to expansion have been described in MSTd.<sup>6,7,8</sup>

We propose that while the expansion neurons could indicate the direction of self-motion when this self-motion is a pure translation (complete overlap between the direction of translation and the direction of gaze), the DDD neurons could indicate the direction of self-motion when the self-motion contains both a translational and a rotational component, i.e., when the direction of translation and the direction of gaze differ. As the angle between gaze and translation changes, these two cell types would provide a continuously changing signal about the direction of translation relative to the object tracked.

Although we propose that the disparity sensitive neurons signal the direction of translation relative to the object fixated, we do not want to imply that disparity is the only signal capable of carrying the depth information needed to compute the direction of self-motion from visual signals. In fact, Warren et al.<sup>2,3,4</sup> have shown that the direction of heading can be determined from visual information in which relative speeds were the only signals about depth. Non-visual information about the self-motion is also available to the subject: such as proprioceptive inputs or corollary discharges about the eye and head movements. In fact, an input about eye movements has been demonstrated in MSTd neurons.<sup>12</sup> It seems reasonable to think that both visual and non-visual information will be used together or separately under different conditions. It was shown recently that Warren and collaborators' conditions<sup>2,3,4</sup> were a special case (very slow tracking eye movements) and that under more general conditions, visual information alone would not be sufficient to provide an accurate signal about translation but that an additional signal about the eye movement was

needed.<sup>13</sup> What we are proposing is that the type of disparity sensitivity described here provides one mechanism that could determine under certain conditions, the observer's direction of motion in the environment by combining, at the single cell level, the relatively low level signals of direction of motion and disparity.

### 5. ACKNOWLEDGEMENTS

This research was done in collaboration with Drs. Robert H. Wurtz and Hidehiko Komatsu at the Laboratory of Sensorimotor Research, National Eye Institute, National Institutes of Health, Bethesda, Maryland. The author was supported by a Centennial Fellowship from the Medical Research Council of Canada.

### 6. REFERENCES

1. J.J. Gibson, The Perception of the Visual World, Houghton Mifflin, Boston, 1950.
2. W.H. Warren and D.J. Hannon, "Direction of self-motion is perceived from optical flow", *Nature*, Vol 336, pp 162 - 163, 1988.
3. W.H. Warren, M.W. Morris & M. Kalish, "Perception of translational heading from optic flow", *J. Exp. Psychol.[Hum. Percept.]*, 14, pp 646 - 660, 1988.
4. W.H. Warren & D.J. Hannon, "Eye movements and optical flow", *J. Opt. Soc. Am. A.*, 7, pp 160 - 169, 1990.
5. J. Rieger and D.T. Lawton, "Processing differential image motion", *J. Opt. Soc. Am. A* 2, pp 354 - 360, 1985.
6. K. Tanaka, K. Hikosaka, H.-A. Saito, M. Yukie, Y. Fukada & E. Iwai, "Analysis of local and wide-field movements in the superior temporal visual areas of the macaque monkey", *J. Neurosci.*, 6, pp 134 - 144, 1986.
7. K. Tanaka & H. Saito, "Analysis of motion in the visual field by direction, expansion/contraction, and rotation cells clustered in the dorsal part of the medial superior temporal area of the macaque monkey", *J. Neurophysiol.*, 62, pp 626 - 641, 1989.
8. C.J. Duffy & R.H. Wurtz, "Sensitivity of MST neuron to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli", *J. Neurophys.*, 65, 1329 - 1345, 1991.
9. C. Wheastone, "Contributions to the physiology of vision I", *Philos Trans.*, B371 - B394, 1838.
10. G.F. Poggio, B. Fischer, "Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey", *J. Neurophysiol.*, 40, pp 1392 - 1405, 1977.
11. J.-P. Roy, H. Komatsu and R.H. Wurtz, "Disparity sensitivity of neurons in monkey extrastriate area MST", *J. Neurosci* 3, pp 2478 - 2492, 1982.
12. W.T. Newsome, R.H. Wurtz, H. Komatsu, "Relation of cortical areas MT and MST to pursuit eye movements II. Differentiation of retinal from extraretinal inputs", *J. Neurophysiol.*, 60, pp 604 - 620, 1988.
13. C.S. Royden, M.S. Banks and J.A. Crowell, "The perception of heading during eye movements", *Nature*, 360, pp 583 - 585, 1992.
14. J.-P. Roy and R.H. Wurtz, "The role of disparity-sensitive cortical neurons in signaling the direction of self-motion", *Nature*, 348, pp 160 - 162, 1988.

## Topography of excitatory and inhibitory connectional anatomy in monkey visual cortex

J. S. Lund, J. B. Levitt and Q. Wu

Department of Visual Science, Institute of Ophthalmology,  
University of London, 11-43 Bath St., London EC1V 9EL, U. K.

### ABSTRACT

It is chiefly within the superficial layers 1-3 of the cerebral cortex that new properties are developed from relayed afferent information. The intrinsic circuitry of these layers is uniquely structured compared to the deeper layers; each pyramidal neuron connects laterally to other pyramids at a series of offset points spaced at regular intervals around it. As seen in tangential sections of layers 1-3, the pyramidal neuron axon terminal fields are roughly circular in cross section, forming a "polka dot" overall pattern of terminal distribution. In regions of peak density, the diameter of the circular fields matches the width of the uninnervated regions between the terminal fields. This dimension is also that of the average lateral spread of the dendrites of single pyramidal neurons making up the connections in each visual cortical area, a dimension which varies considerably between different cortical regions. Since every point across each cortical area shows similar laterally spreading patterns of connectivity, the overall array is believed to be a continuum of offset connectional lattices. It is also presumed that each pyramidal neuron, as well as projecting to separate points, receives convergent inputs from similar arrays of offset neurons.

The geometry of local circuit inhibitory neurons matches elements of these lattices; basket neuron axons in these layers spread three times the diameter of the local pyramidal neuron dendritic fields while the basket neuron dendritic field matches that of the pyramidal cell. If both basket cell and pyramidal neuron at single points are coactivated by afferent relays, the basket axon might create a surround zone of inhibition preventing other pyramidal cells in the surrounding region being active simultaneously. As the pyramid develops its connections this inhibitory field may force each pyramidal neuron to send its axon out beyond the local inhibitory zone to find other pyramidal cells activated by the same stimulus. Since the basket neuron also contacts other basket neurons<sup>18</sup>, by disinhibition through offset basket neurons, it will simultaneously encourage activity in pyramidal cells in a zone outside the limit of its axon field. This scaling of basket neuron axons is present in early postnatal cortex and it could lead to the punctate patterns of pyramidal neuron connectivity which also appear to develop postnatally<sup>24</sup>. This anatomy might also produce the regular spacing of different functional attributes that is typical of visual cortical organization.

Models that explore spatial geometries of excitation and inhibition resembling those described above are urgently needed to test current biological hypotheses underlying investigations of cerebral cortex.

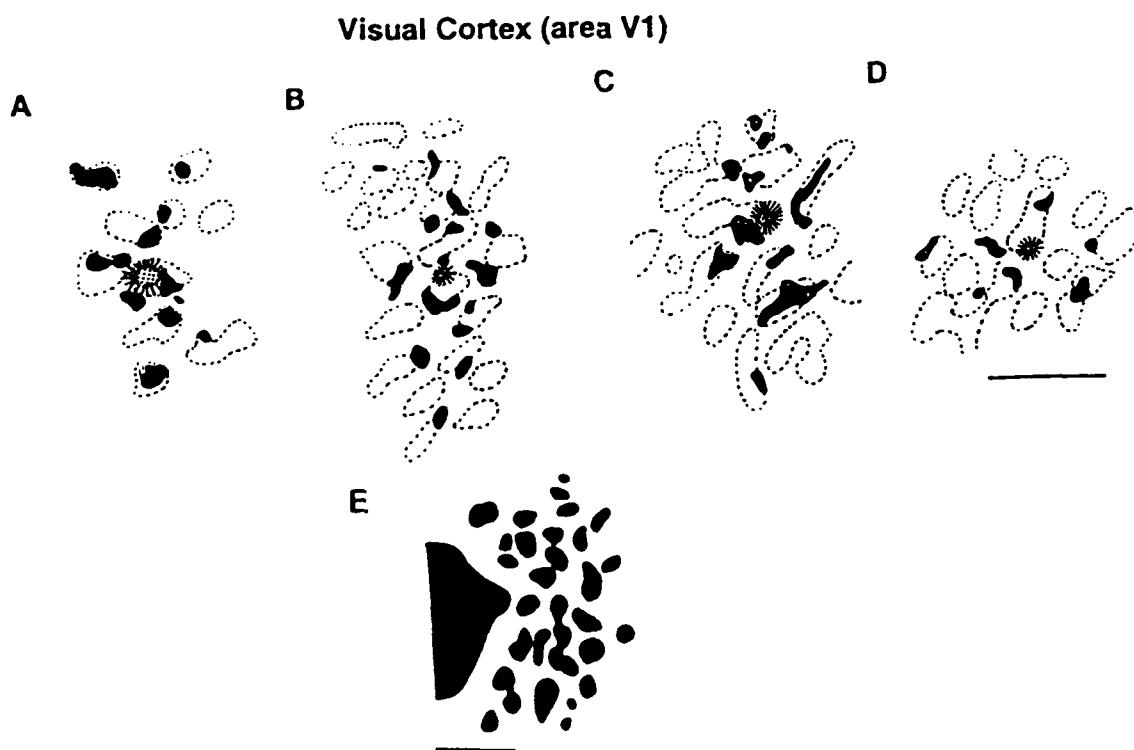
### 1. INTRODUCTION

In the field of vision research there have been active interactions between investigators studying biological forms of vision and those attempting to construct artificial or machine vision. For the purpose of trying to understand how the biological visual system works, theorists have been constructing network models simulating various aspects of visual information processing. These models, some of which have been based on neuroanatomical findings, have attempted to simulate neurophysiological data<sup>2, 11</sup>. However, to a neurobiologist investigating the microcircuitry of the visual system, even some very sophisticated models appear over simplified compared to real neural networks and yet they often ignore simple and obvious details of the real anatomy. We believe it is of the greatest importance that modellers make use of the vast database concerning the anatomy of the visual system obtained by neurobiologists using various anatomical tracing techniques and methods of mapping the topography of functionally active neuron groups. We believe this information is of vital importance to those modellers who aim at biologically realistic simulations or who are attempting to find new neural network architectures. For the biologist there is an urgent need for models based on specific and detailed neuroanatomical findings to aid them in the interpretation of existing data and for guiding further experimental work. In this article we will concentrate on one particularly prominent anatomical feature of the visual cortex, intraareal long distance patterned connections, and we will present our intuitive models for the anatomical findings. We hope that our discussion will call the attention of the computational vision community to this general and interesting feature of cortical organization. We think it likely that ideas coming from anatomical investigations of the biological visual system could also be useful in suggesting computer architectures for solving vision problems<sup>1</sup>.

## 2. GEOMETRIES OF CONNECTIVITY IN PRIMARY VISUAL CORTEX

Anatomical studies of the cerebral cortex in macaque monkeys and other mammalian species have revealed striking geometric arrangements of afferent fiber terminations, intrinsic connectivity and efferent cell grouping. These geometries include stripe-like alternation of two or more functionally different afferents or of efferent cell distribution across single cortical areas, discontinuous "polka dot" patchy distribution of afferents or efferent neurons, and stripe-like or patchy distribution patterns of intrinsic connections. It becomes an important question whether these geometries reflect a connectional strategy essential to generating new functions in the region, or if the geometry may be an event that is of no functional importance beyond perhaps establishing interdigitated spatially coherent "maps" of the sensory or motor periphery during development.

In our recent work on the anatomy and topography of both excitatory and inhibitory intrinsic connections in the macaque monkey visual cortical areas<sup>25, 26, 27, 44</sup>; we have begun to wonder if the topography of the intrinsic connections may be an essential determinant of function; moreover, if network models based on these patterns were to have interesting properties it might be possible to use such models in a predictive fashion to guide further anatomical and physiological studies of cortical organization. While the anatomical information we will discuss in this article is incomplete and its relation to function is not understood, the information may prompt other workers to devise theoretical models (as was begun by Mitchison and Crick<sup>31</sup>) which could clarify ideas concerning the function of such connectivity and suggest essential features to look for in anatomy and function of the circuits based on features that appear to constrain the models.



**Figure 1.** Small iontophoretic injections of biocytin (fringed stippled circles) made into macaque area V1. Biocytin-labeled terminal patches (black) arising from the lateral projections of local pyramidal neurons are reconstructed in relation to CO-rich "blobs" which are sites of thalamic axon terminations (with borders marked by broken lines). In A, a "blob" injection gives labeling mainly in blobs. B-D are injections on the edges of CO-rich blobs, and their projections are mainly but not exclusively to "edge" positions. Note that the intrinsic system of orthograde terminal patches can much smaller than the area of single blobs. Also note that injections into interblob zones would produce patches of termination mainly in interblob regions. Scale bar for A-D, 1mm. E. Map of patchy terminal label in layer 2-3 of macaque V1 to one side of a very large pressure injection of biocytin (shown to left of the figure). Note that distinct patches of terminal persist without extensive fusion. Scale bar for E, 1 mm. (From Lund et al. <sup>26</sup>)

A primary function of the superficial layers of the primary visual cortex (V1) appears to be the generation of new properties from relayed afferent information. The intrinsic circuitry of these layers is uniquely structured<sup>32</sup> compared to the deeper layers; each pyramidal neuron, or some significant and evenly distributed proportion of the total population of pyramids, connects via long distance axon projections to other pyramids<sup>14, 30, 32</sup> offset laterally at a series of points spaced at regular intervals around it (see Figure 1). As seen in histological sections cut parallel to the pia, pyramidal neuron axon terminal fields labeled from a single, small but intense iontophoretic injection of biocytin within layers 2-3 are approximately circular in cross section, forming an overall "polka dot" pattern of evenly spaced regions of termination<sup>26, 42, 44</sup>.

The maximum diameter of the fully labeled patches of terminals matches the width of the uninnervated gaps between them; if the size of the injected point is increased substantially the patchy system of connectivity is still evident and the terminal patch size and width of gaps between them does not increase appreciably or fill in (see Figure 1E). The dimension of the patch and gap widths is interesting; it is very closely matched to the diameter of the dendritic field spread of single pyramidal neurons giving rise to the system of connectivity; this dimension is approximately 240 $\mu$ m in the primary visual cortex<sup>26</sup>. Since any point injected across the superficial layers of area V1 shows a similar lattice-like pattern of connectivity it is presumed that the system is a continuum of offset connectional lattices. The total field of connected points around single small injections extends far enough laterally in the superficial layers of V1 to link together cells that share some part of their functional minimum response fields to visual stimuli with neurons at the injected point<sup>44</sup>. The cross correlation studies by Ts'o and his colleagues<sup>38</sup> show that in area V1 these connections predominantly link clusters of cells with common properties; in addition some smaller proportion of the connections appear to link together regions that may differ in at least some properties, for example linking regions of opposite ocular dominance or linking regions with and without input from the intercalated layers of the LGN<sup>44</sup>. It is possible to show these connections with both anterograde and retrograde tracer substances and it is believed that each pyramidal neuron, as well as projecting to separate points, also receives convergent input from similar arrays of offset neurons.

### 3. COMPARISON BETWEEN AREA V1 AND OTHER AREAS

Our anatomical studies on visual association cortex<sup>19, 43</sup> show that areas V2 and V4 have similar connectional lattices to area V1 in their superficial layers. There are however some differences in scale and overall pattern from the connections seen in area V1. First, within either V2 or V4, the terminal patch and intervening gap size are of equal dimensions, as was seen in the V1 intrinsic connections; however, the absolute size of these elements is larger in V2 than in V1, and larger in V4 than in either of the other two areas. Interestingly, when the mean size of the pyramidal neuron dendritic fields was examined in areas V2 and V4 they also exceeded the size of the V1 neurons, especially in area V4, and each was a reasonably good match to the dimension of the patch and gap in their own area. We went on to examine primary somatosensory, motor and prefrontal cortical areas and found that each area had superficial layer lattice connections, that the gap and terminal patch sizes were matched in width in each area and that the size of single pyramidal neuron dendritic field spread in the lattices closely matched the elements of the lattice repeat in each area, despite a twofold difference in size between the largest and smallest lattice dimensions observed in different areas<sup>20, 26</sup>. These common features between different cortical areas in the intrinsic connectivity of their superficial layers is not just a feature of the primate; very similar patterns of connectivity are seen in at least the visual cortices of other species (cat and tree shrew) and the same feature of a similar dimension to terminal patch, uninnervated gap and pyramidal neuron dendritic field width is common to each (see Figure 2).

Some additional features to the organization of the lattice arrays as seen in area V1 are found in area V2 (see Figure 3). The territory of area V2 is divided into three compartments laid out in interleaved parallel stripe-like arrays<sup>23</sup>; these compartments are distinguished by receiving functionally different sets of afferents from area V1. The geometries of the superficial layer lattice array of pyramidal neuron connections in V2 obey the constraint of equal gap and patch width, with match to diameter of single pyramidal system dendritic fields, in their connections within single stripe compartments. However, even small deposits of biocytin limited to single compartments produce axon projections which extend to give terminal patches in all three compartments and these projections to other stripe compartments often require exceptionally long axon trajectories without terminals to reach their destination, suggesting that axon geometries between stripes are decided by different factors than the within stripe connections. If the long axons make several patches of connections within a single stripe, even if that stripe is not the same kind of compartment as the injection site, the patches of terminals again obey the geometry of a within stripe system.

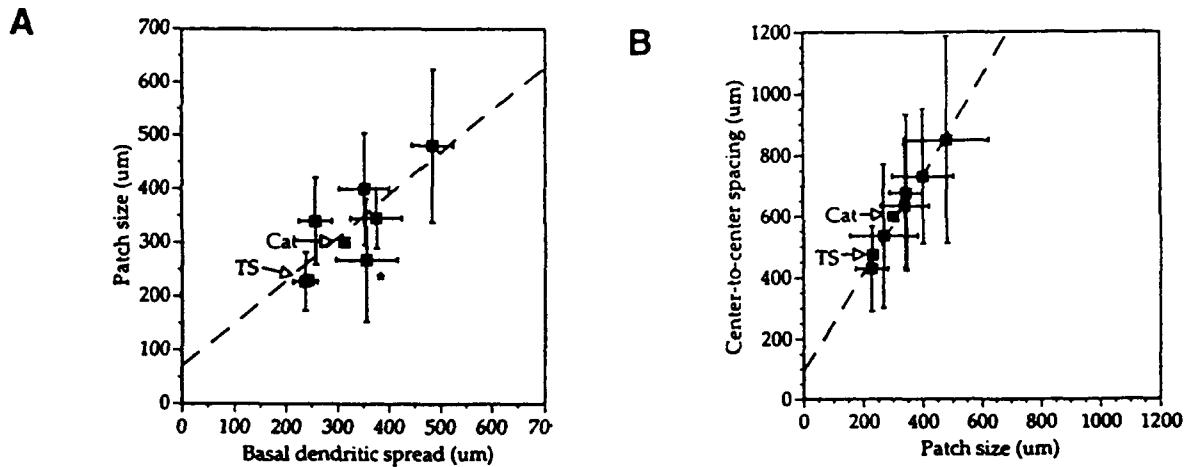


Figure 2. A plot of the relationship between the average diameter of terminal patches (or width of stripe-like zones) and the average lateral spread of the basal dendritic field of single layer 2-3 pyramidal neurons. Error bars indicate SDs of the data points. The dashed line is a least-squares regression line through the monkey data. Prefrontal data point (elongated stripe-like terminal zones rather than patches) is marked by asterisk. There is a significant correlation between these measures ( $r = 0.779$ ,  $p < 0.05$ ), and the data fall very near to a line of slope 1, which indicates that the size of these terminal zones is scaled almost precisely to the size of the local pyramidal neuron basal dendritic field. We have also included data from area V1 of cat and tree shrew (TS) (indicated by arrows) to illustrate that similar constraints seem to hold in these non-primate species as well. B. Plots of terminal patch center-to-center spacing against patch size for the same macaque cortical areas. The least-squares regression line through the monkey data is indicated by the dashed line. These measures were highly correlated ( $r = 0.989$ ,  $p < 0.001$ ), indicating that the spacing of these terminal zones across the cortex is matched to the size of the patches themselves, thus maintaining nearly equivalent coverage by the lattice in each cortical area. Data from cat and tree shrew (indicated by arrows) are again plotted for comparison. (From Lund et al.<sup>26</sup>)

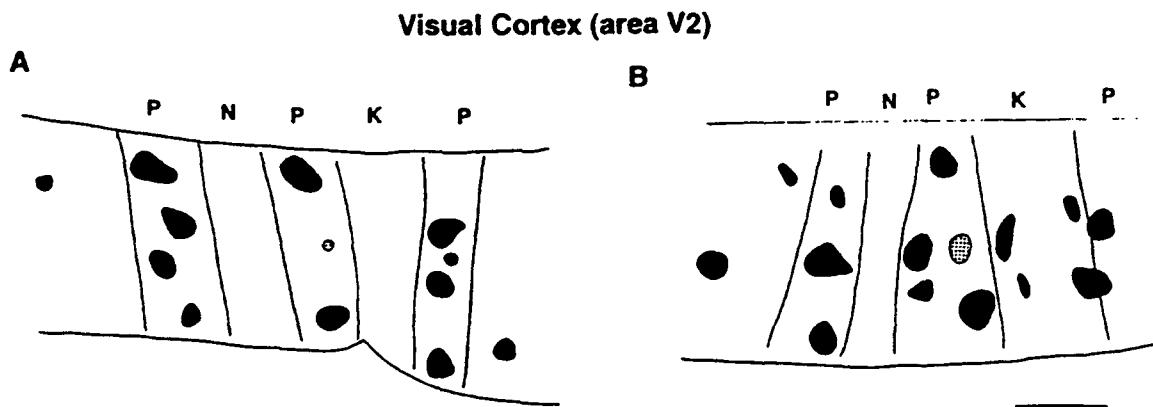


Figure 3. Reconstructed maps of terminal labels following two (A, B) iontophoretic biocytin injections into area V2 in the lunate sulcus. Injection sites are indicated by the stippled circles; terminal label, by the black regions. V2 has been flattened and sectioned tangentially; the solid lines indicate the borders of the CO stripe compartments (K: thick; N: thin; P: pale). Both of these injections were made in pale stripes, and the most (though not all) terminal label was also in pale stripes. Scale bar, 1mm. (From Lund et al.<sup>26</sup>)

#### 4. DEVELOPMENT OF CONNECTIONAL LATTICES

We suggest from these findings that there is an innate propensity for the superficial cortex to develop repetitive lattice-like connectivity and that this feature is not a special feature of any particular sensory modality or species. We would like to know if this connectivity confers any special properties on the function of the neuropil and what constrains its development. Studies examining the visual cortex during development report an absence of the lattice connectivity in the early postnatal period<sup>10</sup> but the existence of long laterally spreading pyramidal neuron axon trunks. As visual stimulation begins to drive cortical activity the patchy pattern of connectivity appears in the form of axon sprouts arising along the long horizontal trunks; it is evidently constrained by the patterns of activity in the neurons since the connections can be driven into monocular domains if the animal is reared with alternating monocular vision<sup>24</sup> or with strabismus<sup>9</sup>, but the repetitive geometry remains unaltered.

#### 5. GEOMETRY OF LOCAL INHIBITION

A constraint that might be operating to create the regular repeating distance is the anatomical geometry of local inhibitory neurons. We have noted that the axons of one type of inhibitory neurons, the GABAergic basket neurons, present at birth in layer 3 of the superficial layers where the periodic patterns of connectivity begin to appear, spread three times the width of the pyramidal neuron basal dendritic arbor. It is known<sup>36</sup> that these axons preferentially contact the somata and proximal dendritic segments of pyramidal neurons. If the basket neuron, whose dendritic field matches that of the local pyramidal neurons in its lateral spread, is driven by the same afferents as the pyramidal neurons local to it, the axon of the basket neuron will create a zone of inhibition around itself. If one assumes a Hebbian rule for development of connections between pyramidal neurons, this zone of inhibition will force colocalized pyramidal neurons to send their axon connections outside this

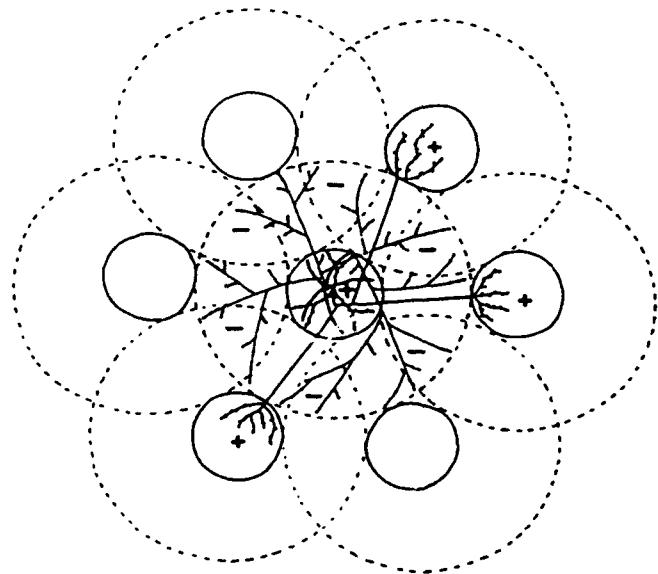


Figure 4. A diagram of intraareal cortical connectivity suggested to explain offset patch-like distribution of pyramidal neuron axon terminals. The cortex is viewed from the surface, and one pyramidal neuron is indicated spatially colocalized with an inhibitory "basket" neuron. The basket neuron axon spreads over a region limited by the innermost hatched circle (indicated by minus sign). Coactivation of both pyramid and basket neurons would drive inhibition within the inner hatched circle, thereby making it less likely that the pyramidal neuron would find any other simultaneously active pyramidal neurons with that region, and thus retarding the establishment of synaptic connectivity between different pyramids in this zone during development. The pyramidal neuron makes connections to zones (small circles marked by plus signs) outside the range of the coactive inhibitory axon field. However, here again, every other pyramidal neuron it contacts will have a similar inhibitory surround from basket neurons colocalized with them (outer dashed circles), and the excitatory connectivity is restricted to a series of six points, in hexagonal form across the cortex. These constraints are the same for any point, so the same hexagonal connectional matrix would be found around any single pyramidal cell, thus forming a continuum across the cortex. (From Lund et al.<sup>26</sup>)

inhibitory field to points where there may be other pyramidal neurons driven by the same stimulus and simultaneously active. Since the basket neuron axon also contacts other basket neurons<sup>18</sup>, it will simultaneously encourage activity in pyramidal cells in a zone outside the limit of its axon field through disinhibition of their basket axon contacts. Figure 4 (from Lund et al.<sup>26</sup>) illustrates this scaling of basket neuron and pyramidal neuron axon connectional patterns. As the pyramidal neuron begins to build connections with neighboring cells, each of the pyramidal neurons it contacts will also have basket neurons local to it; since the pyramidal neuron axons also contact smooth dendritic, presumed GABAergic neurons, it is possible that they, like the afferents, also drive the local basket neurons and this will encourage the connections between the pyramidal neuron axon connections to make regularly sized steps in a roughly hexagonal array.

## 6. TOPOGRAPHY OF FUNCTION

The topography of inhibition, if active as outlined above, could be an essential element in determining the topography of function in the superficial layers. It is clear in visual cortical area V1 that particular functional attributes, such as specificity for a particular orientation of line, repeat at regularly spaced intervals across the superficial cortex; the mean repeat distance for cells having similar responses for any one function is close to (or perhaps slightly larger than) the center-to-center distance between adjacent patches of the intrinsic connectional lattice system, and the distance between opposite extremes, e.g. orthogonal orientations, is half that distance. However, functions in the superficial layers appear to be distributed across the cortical sheet as gradually changing parameters between extreme values (e.g. gradual change in orientation specificity) rather than having sharp boundaries with a sudden change between one function and another. But, because it is impossible to have a smooth change in function in all directions across a two dimensional sheet, occasional abrupt changes or breaks in function are observed, both in recording experiments<sup>16, 17</sup> and in functional imaging<sup>3, 5</sup>. The connectional lattices may not exactly match the repeat of single functions in the spacing between patches since the lattice may connect those points at which several functions, not just orientation, are well correlated.

As illustrated in Figure 4, the geometry of excitatory and inhibitory zones of pyramidal and basket neurons in the superficial layers is likely to enforce the cortical sheet to be organized into an array of hexagons. Based on theoretical considerations, some researchers<sup>6, 35, 40, 41</sup> have suggested that orientation columns may well be organized into hexagons (corresponding to hypercolumns) so that the whole cortical sheet is an array of hexagons. The cortex has been faced with the problem of how to compact and interrelate many two-dimensional maps in a single cortical sheet. Swindale<sup>37</sup> has discussed in theoretical form such an issue, calling it a "dimension reduction" problem, and suggested that it is closely analogous to the traveling salesman problem well studied by theorists.

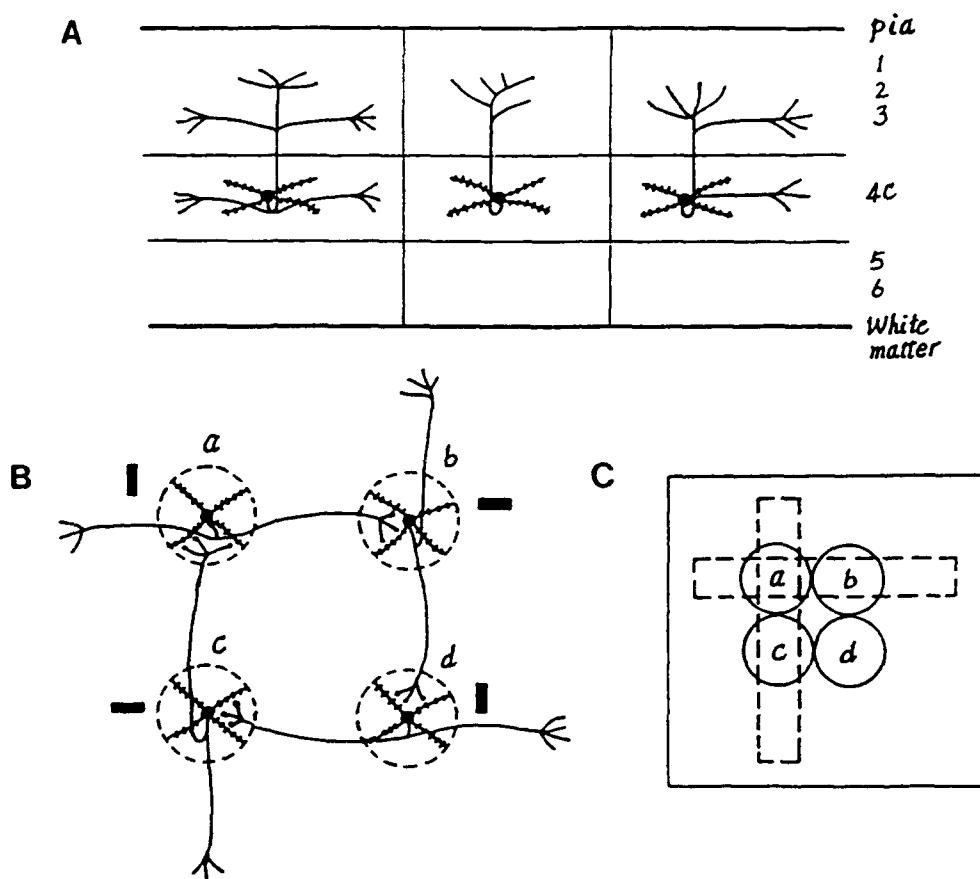
## 7. DISTRIBUTION OF AFFERENTS

While the terminals of afferents to the superficial layers of the primary visual cortex occupy non-overlapping territories (thalamic inputs to the cytochrome oxidase (CO) rich "blobs"<sup>13</sup> and spiny stellate neuron projections from mid layer 4C to CO-poor interblob regions)<sup>44</sup>, the dendrites of the pyramids lap freely across the junctions between these different afferent territories<sup>28</sup>. Because single afferent territories are repeated at the same scale as the intrinsic system connectional territories, and therefore matched also to the dimensions of single pyramidal neurons, only at the very center of each afferent territory, e.g. at the center of a CO-rich blob or at the center of an interblob region, will there be a small population of pyramidal neurons whose dendrites are totally within that compartment. Functions expressed most clearly at the center of blobs (e.g. monocularity, no orientation specificity) gradually change between blob and interblob territories to become binocular and specifically tuned to orientation. The distribution of afferents to separate compartments may be determined by the same substrate of inhibition as that shaping the laterally spreading lattice connections since they are scaled to the same dimension (see Figure 1); however, while the intrinsic lattice system is continuously distributed, presumably relating to individual cell properties resulting from the sum of inputs to each neuron, the terminal distribution for each population of afferents is discontinuously distributed much as is seen for the populations of right and left eye afferents in the underlying layer 4C. This difference may be due to the afferent fiber populations having such markedly different response properties that they tend to segregate on the continuum of pyramidal neuron dendritic surface to exclusive territories, whereas the intrinsic lattice system collaterals, while just as discretely parcellated, reflect in their continuum the pyramidal neuron postsynaptic responses determined by gradients of overlap of their dendrites into different pools of afferent fibers.

## 8. ORIGIN OF ORIENTATION SPECIFICITY GRADIENTS

One of the leading questions in the organization of the visual cerebral cortex has been the question of what the substrates may be for determining orientation specificity of its constituent neurons. We have raised the possibility here that the smooth gradients of functional change, including that of change in orientation specificity, seen to occur across the superficial cortical

sheet are the result of the geometries of inhibition and dendritic overlap between parcellated afferent zones. How then could orientation specificity be built from such a model? It could require that two sets of afferents enter the superficial neuropil and have sufficiently different response properties that they establish non-overlapping terminal territories; if the difference in functional properties of the two populations were responses to orthogonal line orientations, then the continuum of overlap of pyramidal neuron dendrites in the superficial layers into these two pools of afferents might be sufficient to build a continuum of response specificities to orientations of line between the two orthogonal extremes. It is presumed that these two sets of afferents would both enter the CO-poor interblob territories since such territories are devoted to the gradual changes in all orientations. We have recently found this territory to receive its principal afferents from neurons positioned in middle depth of layer 4C<sup>44</sup> and certain features of the axon projections of these neurons suggest how they may develop into sets with orthogonal line orientation preferences.



**Figure 5.** A. Anatomical projections hypothesized for spiny stellate neurons of layer 4C of macaque monkey visual cortex, based on observed projections of small clusters of neurons labeled by small injections of biocytin into layer 4C. B. Suggestion for configuration of lateral connections in layer 4C that would produce orthogonal orientation specificities. Layer 4C is viewed from above and the lateral connections of four spiny stellate neurons (a - d) are indicated. It should be noted that the connections between layer 4C neurons are hypothesized to exert a subliminal effect that enhances responses to thalamic inputs in the postsynaptic neurons but do not themselves drive the postsynaptic cells. C. Visual field positions of four spiny stellate neurons in B (a - d). Hatched bars indicate line positions that would produce horizontal orientation preference in neuron b and vertical orientation preference in neuron a.

Small injections of biocytin made into layer 4C can produce several different patterns of local axon spread within the same 4C layer as well as different patterns of termination in layer 3 immediately above the injection site. When viewed in single sections cut from pia to white matter, axon collaterals may spread only very locally to the injection site in layer 4C and rising trunks passing to layer 3B will give a simple fan-like arbor whose only eccentricity will be an avoidance of blob territories if they lie immediately above the injection site. Another pattern seen includes additional axon collaterals passing along the length of a spur of neuropil or to a laterally offset point in mid layer 4C 300-500 $\mu$ m from the injection site, either just to one side of the injection or to both sides (see Figure 5A). These sidesteps are often accompanied by a similar sized overlying lateral sidestep off the rising axon trunks terminating in layer 3B immediately over the position of the layer 4C sidestep terminations. Small biocytin injections in CO-poor interblob regions of layer 3B always retrogradely label a cluster of neurons immediately below the injection site; in addition they often also retrogradely label small clusters of neurons offset laterally from the injection axis in mid layer 4C, confirming the picture seen from orthograde labeling<sup>45</sup>.

Further anatomical studies are needed to carefully map the geometry of the layer 4C lateral projections but Figure 5B&C suggests one possible configuration of the connectivity that could produce two pools of neurons with orthogonal orientation preferences that could present the extremes for the smoothly changing sequences of orientation specificity seen across the neuropil of the superficial layers. Figure 5B illustrates these projections in a tangential view of layer 4C, and suggests how these sidestepping connections could generate orthogonal orientation preferences. This is achieved by linking neurons with adjacent circular receptive fields across two opposite axes of the precise retinal map in layer 4C. These connections would produce an enhanced response in the postsynaptic neuron by virtue of the additional excitatory reinforcement of its excitatory reaction when lines are oriented such that they cross the receptive fields of both pre- and post-synaptic neurons simultaneously. Physiological recording of neurons within layer 4C shows the existence of neurons with orientation preferences<sup>4, 15, 21</sup>. The model would presume a tight map of enhanced responses to orthogonal orientations (i.e., orientation preferences) within mid to upper layer 4C, while at the same time the layer 4C cells should still maintain brisk responses to any stimulus orientation.

## 9. DISTRIBUTION OF EFFERENT NEURON POPULATIONS

If the intrinsic lattice-like connections link together neurons of particular response properties, it would be logical to expect that connected neuron pools would then serve as the sets of efferent neurons presenting information to the next destination. In area V1 the lattice connections seem to link every point across the cortical surface so it comes as somewhat of a surprise to find that two different efferent neuron pools of the superficial layers are restricted to non-overlapping territories defined by different afferent fiber terminal zones: the CO-rich blobs (which receive input from the intercalated zones of the dorsal lateral geniculate nucleus and project to the "thin" CO-rich stripes of area V2) and the CO-poor interblob zones (which receive input from mid layer 4C and project to the CO-poor "pale" stripes of V2)<sup>21, 23</sup>. It is an equal surprise to find that the distribution of different efferent cells of area V2 are also very closely linked to the afferent stripe-like compartments in the region<sup>34, 39, 46</sup> when labeling of the intrinsic lattices clearly shows any single point across the cortex to distribute axon connections widely to all the three compartments.

While the subtleties of individual pyramidal neuron dendritic sampling from several sets of interleaved or intermingled repetitive gradient properties across the V1 cortical sheet may enable activation in the superficial layers of V1 of a unique lattice representation of the combined qualities of any particular stimulus, it is unlikely that the activity of every possible permutation of lattice activity be represented in the output from the V1 region. Rather, the output to the next stage in visual processing could once more combine two extremes for each of a limited number of functions and these extremes could form the basis for extrapolations and development of new functional properties in the next region. Since it is true that neurons of either blob and interblob compartments in V1 make the majority of their connections within the same compartment<sup>43, 44</sup> and that these two compartments contain the entire population of neurons in layers 2-3, it is possible that outputs from neurons restricted to either one of these compartments could have functional properties representing the extremes of newly developed properties (e.g., color contrast or three dimensional shape discrimination). These extremes may then map with segregated afferent terminal zones within single stripe compartments in V2 as the basis for constructing further gradient functions.

## 10. DIFFERENT GEOMETRIES OF LATTICES AND EXCEPTIONAL SPECIES

As is usual with all biological phenomena, exceptions have been found to the general prevalence of punctate lattice connectivity in the superficial layers of the mammalian cortex. These exceptions may be particularly important in providing clues to features essential to development of punctate lattice connectivity. One difference is that instead of the lattice terminal fields having a "polka dot" geometry they can terminate with a stripe-like geometry with equal width of terminal stripes and interleaved gaps of uninervated neuropil. This geometry has been described for the primary visual cortex of the tree shrew<sup>33</sup>.

and also in the macaque monkey prefrontal cortex<sup>20, 26</sup>. Stripe and gap width are again similar to that of the dendritic field spread of single pyramidal neurons of the same layers. We have suggested<sup>26</sup> that the stripe geometry could be produced by inhibition, much as suggested above for the more common punctate arrays, if the basket neuron axons have an elongated slab-like form instead of a circular geometry (see Figure 6). We have not yet traced the three dimensional geometry of the prefrontal cortex or tree shrew basket neurons to check their orientation but it is known that such neurons can have anisotropic distribution of their axon arbors<sup>29</sup>.

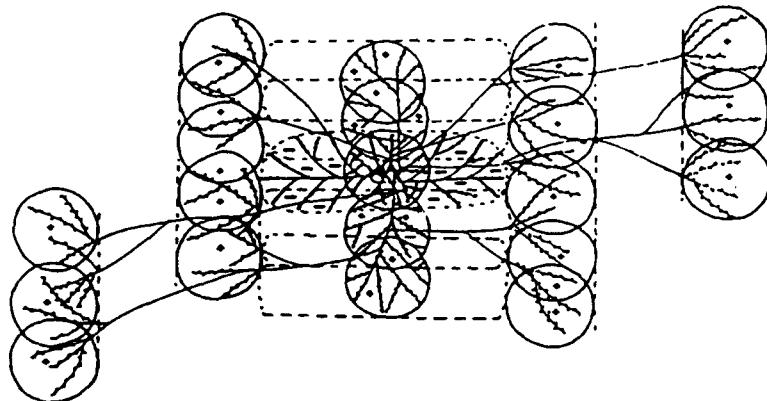


Figure 6. A diagram of cortical connectivity suggested to explain stripe-like discontinuous distribution of intraareal pyramidal neuron connectivity in macaque prefrontal cortex. Here the basket neuron colocalized with the pyramidal neuron provides an elongated inhibitory area field (marked by minus signs); this permits the colocalized and coactivated pyramidal neuron to make local excitatory connections to simultaneously active pyramids in a band orthogonal to the long axis of the basket neuron inhibitory axon field. The presence of the inhibitory field vetoes synaptic connections being made within the flanking stripes and the pyramidal neuron axon "steps" over a stripe-like region before establishing more distant terminal fields—as in Figure 4, here again under the same constraint so that stripes of terminal label continue across the cortex. (From Lund et al.<sup>26</sup>)

Another major exception to the general presence of connectional lattices is the rat cortex where, so far, we have been unable to demonstrate the existence of intraareal lattices despite excellent transport of biocytin locally and interareally. Burkhalter<sup>7</sup> and Burkhalter and Charles<sup>8</sup> using HRP and *Phaseolus* lectin as tracer substances have reported a waxing and waning of intrinsic connectivity in the rat visual cortex but pointed out that this pattern does not resemble the isolated patches or stripes seen in other species. In the rat these fluctuating connections appear to arise from layer 5 and it is noticeable that layers 2-3 in the rat visual cortex occupy a much smaller proportion of the total cortical depth than layers 2-3 in monkeys, cats and tree shrews; functionally, the mouse visual cortex, at least, lacks the gradually changing sequences of orientation specific cells seen in other species, although the cells do not lack orientation specificity<sup>12</sup>. The rat cortex may also lack basket neurons with wide spreading axons (Somogyi-personal communication). Therefore, it is possible that the rodent cortex has a different basic pattern of organization than the other species discussed and it will be worth examining their cortical organization for further differences in structure and function. These differences in the rodent again raise the question of the functional import of the lattices seen in primates and other species.

## 11. SUMMARY

The hypotheses discussed above, concerning the way observed anatomical connectional geometries may determine the development of functional characteristics within the primary visual cortex, need to be examined in detail by theorists to determine if they are feasible and worth further investigation. There are many issues that need to be considered at both single cell level (for example, the feasibility that pyramidal neurons compute a mean value between two sets of inputs of opposite specificity simply from the numerical weights of synapses they receive from each source) and at system level (for example, why the different efferent neuron sets do not distribute according to the linkage of neurons seen in the intrinsic lattice systems but, instead, revert to the distribution of the afferent territories to blob and interblob zones of area VI and to stripe-like territories in V2). Predictions from the models are of the utmost importance since they will force neurobiologists to search for

information that may confirm or refute particular hypotheses for cortical functional anatomy and in this way quicken and guide the progress of further cortical studies.

## 12. ACKNOWLEDGMENTS

This research was supported by MRC grant G9203679N and NIH-NEI grant EY 10021.

## 13. REFERENCES

1. Y. Aloimonos and A. Rosenfeld, "Computer vision", *Science*, Vol. 253, pp. 1249-1254, 1991.
2. M. A. Arbib and A. R. Hanson (Eds.), *Vision, Brain, and Cooperative Computation*, MIT Press, Cambridge, MA, 1987.
3. G. G. Blasdel, "Orientation selectivity, preference and connectivity in monkey striate cortex," *J. Neurosci.*, Vol. 12, pp. 3141-3163, 1992.
4. G. G. Blasdel and D. Fitzpatrick, "Physiological organization of layer 4 in macaque striate cortex," *J. Neurosci.*, Vol. 4, pp. 880-895, 1984.
5. E. Bartfeld and A. Grinvald, "Relationships between orientation preference pinwheels, cytochrome oxidase blobs, and ocular-dominance columns in primate striate cortex," *Proc. Natl. Acad. Sci. USA*, Vol. 89, pp. 11905-11909, 1992.
6. V. Braitenberg and C. Braitenberg, "Geometry of orientation columns in the visual cortex", *Biol. Cybern.*, Vol. 33, pp. 179-186, 1979.
7. A. Burkhalter, "Intrinsic connections of rat primary visual cortex: Laminar organization of axonal connections," *J. Comp. Neurol.*, Vol. 279, pp. 171-186, 1989.
8. A. Burkhalter and V. Charles, "Organization of local axon collaterals of efferent projection neurons in rat visual cortex," *J. Comp. Neurol.*, Vol. 302, pp. 920-934, 1990.
9. A. Burkhalter and L. Tychsen, "Alteration of intracortical connections in monkeys who lacked normal binocular experience in infancy," *Investig. Ophthal. Vis. Sci. ARVO Abst.*, Vol. 34, p. 1173, 1993.
10. E. M. Callaway and L. C. Katz, "Emergence and refinement of clustered horizontal connections in cat striate cortex," *J. Neurosci.*, Vol. 10, pp. 1134-1153, 1990.
11. P. S. Churchland and T. J. Sejnowski, *The Computational Brain*, MIT Press, Cambridge, MA, 1992.
12. U. C. Dräger, "Receptive fields of single cells and topography in mouse visual cortex," *J. Comp. Neurol.*, Vol. 160, pp. 269-290, 1975.
13. D. Fitzpatrick, K. Itoh and I. T. Diamond, "The laminar organization of the lateral geniculate body and the striate cortex in the squirrel monkey (*Saimiri Sciurus*)," *J. Neurosci.*, Vol. 3, pp. 673-702, 1983.
14. C. D. Gilbert and T. N. Wiesel, "Clustered intrinsic connections in cat visual cortex," *J. Neurosci.*, Vol. 3, pp. 1116-1133, 1983.
15. M. J. Hawken and A. J. Parker, "Contrast sensitivity and orientation selectivity in lamina IV of the striate cortex of old world monkeys," *Exp. Brain Res.*, Vol. 54, pp. 367-372, 1984.
16. D. H. Hubel and T. N. Wiesel, "Sequence regularity and geometry of orientation columns in the monkey striate cortex," *J. Comp. Neurol.*, Vol. 158, pp. 267-293, 1974.
17. D. H. Hubel and T. N. Wiesel, "Functional architecture of macaque monkey visual cortex," *Proc. Roy. Soc. London B*, Vol. 198, pp. 1-59, 1977.
18. Z. F. Kisvárday, C. Beaulieu and U. T. Eysel, "Networks of GABAergic large basket cells in cat visual cortex (area 18): Implication for lateral disinhibition," *J. Comp. Neurol.*, Vol. 327, pp. 398-415, 1993.
19. J. B. Levitt, T. Yoshioka and J. S. Lund, "Intrinsic connectivity in macaque V2: Evidence for interactions between functional streams," *Soc. Neurosci. Abst.*, Vol. 18, p. 294, 1992.
20. J. B. Levitt, D. A. Lewis, T. Yoshioka and J. S. Lund, "Topography of pyramidal neuron intrinsic laminar connections in macaque monkey prefrontal cortex (areas 9 and 46)," *J. Comp. Neurol.*, in press.
21. M. S. Livingstone and D. H. Hubel, "Anatomy and physiology of a color system in the primate visual cortex," *J. Neurosci.*, Vol. 4, pp. 309-356, 1984.
22. M. S. Livingstone and D. H. Hubel, "Specificity of intrinsic connections in primate primary visual cortex," *J. Neurosci.*, Vol. 4, pp. 2830-2835, 1984.
23. M. S. Livingstone and D. H. Hubel, "Connections between layer 4B of area 17 and the thick cytochrome oxidase stripes of area 18 in the squirrel monkey," *J. Neurosci.*, Vol. 7, pp. 3371-3377, 1987.
24. S. Löwel and W. Singer, "Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity," *Science*, Vol. 255, pp. 209-212, 1992.
25. J. S. Lund and T. Yoshioka, "Local circuit neurons of macaque monkey striate cortex III: Neurons of laminae 4B, 4A, and 3B," *J. Comp. Neurol.*, Vol. 311, pp. 234-258, 1991.

26. J. S. Lund, T. Yoshioka and J. B. Levitt, "Comparison of intrinsic connectivity in different areas of macaque monkey cerebral cortex," *Cerebral Cortex*, Vol. 3, pp. 148-162, 1993.

27. J. S. Lund, T. Yoshioka, and J. B. Levitt, "Substrates for interlaminar connections in area V1 of macaque monkey cerebral cortex," In *Cerebral Cortex 10: Primary Visual Cortex in Primates*, A. A. Peters and K. S. Rockland (Eds.), Plenum Press, New York, in press.

28. R. Malach, "Dendritic sampling across processing streams in monkey striate cortex," *J. Comp. Neurol.*, Vol. 315, pp. 303-312, 1992.

29. M. Marin-Padilla, "Three dimensional reconstruction of the pericellular nests (baskets) of the motor (area 4) and visual (area 17) areas of the human cerebral cortex: A Golgi study," *Z. Anat. Entwicklungs gesch.*, Vol. 144, pp. 123-135, 1974.

30. B. A. McGuire, C. D. Gilbert, P. K. Rivlin and T. N. Wiesel, "Targets of horizontal connections in macaque primary visual cortex," *J. Comp. Neurol.*, Vol. 305, pp. 370-392, 1991.

31. G. Mitchison and F. Crick, "Long axons within the striate cortex: Their distribution, orientation and pattern of connections," *Proc. Natl. Acad. Sci. USA*, Vol. 79, pp. 3661-3665, 1982.

32. K. S. Rockland and J. S. Lund, "Intrinsic laminar lattice connections in primate visual cortex," *J. Comp. Neurol.*, Vol. 216, pp. 303-318, 1983.

33. K. S. Rockland, J. S. Lund and A. L. Humphrey, "Anatomical banding of intrinsic connections in striate cortex of tree shrews," *J. Comp. Neurol.*, Vol. 209, pp. 41-58, 1982.

34. S. Shipp and S. Zeki, "Segregation of pathways leading from area V2 to areas V4 and V5 of macaque monkey visual cortex," *Nature*, Vol. 292, pp. 543-545, 1985.

35. K. V. Shenoy, J. Kaufman, J. V. McGrann and G. L. Shaw, "Learning by selection in the trion model of cortical organization," *Cerebral Cortex*, Vol. 3, pp. 239-248, 1993.

36. P. Somogyi, Z. F. Kisvárday, K. A. C. Martin and D. Whitteridge, "Synaptic connections of morphologically identified and physiologically characterized large basket cells in the striate cortex of cat," *Neuroscience*, Vol. 10, pp. 261-294, 1983.

37. N. V. Swindale, "Elastic nets, traveling salesman, and cortical maps," *Current Biol.*, Vol. 2, pp. 429-431, 1993.

38. D. Y. Ts'o, C. D. Gilbert and T. N. Wiesel, "Relationships between horizontal connections and functional architecture in cat striate cortex as revealed by cross-correlation analysis," *J. Neurosci.*, Vol. 6, pp. 1160-1170, 1986.

39. D. C. Van Essen, D. J. Feilleman, E. A. DeYoe, J. Olavarria and J. Knierim, "Modular and hierarchical organization of extrastriate visual cortex in the macaque monkey," *Cold Spring Harbor Symp. Quant. Biol.*, Vol. LV, pp. 679-696.

40. W. Von Seelen, "Zur Informationsverarbeitung im visuellen System der Wirbeltiere," *Kybernetik*, Vol. 7, pp. 89-109, 1970.

41. A. B. Watson, "Architecture of visual cortex", in *Vision: Coding and Efficiency*, C. Blakemore (Ed.), Oxford University Press, Oxford, 1990.

42. T. Yoshioka, G. G. Blasdel, J. B. Levitt and J. S. Lund, "Patterns of lateral connections in macaque visual area V1 revealed by biocytin histochemistry and functional imaging," *Soc. Neurosci. Abst.*, Vol. 18, p. 299, 1992.

43. T. Yoshioka, J. B. Levitt and J. S. Lund, "Intrinsic lattice connections of macaque monkey visual cortical area V4," *J. Neurosci.*, Vol. 12, pp. 2785-2802, 1992.

44. T. Yoshioka, J. B. Levitt and J. S. Lund, "Anatomical basis for channel interactions in macaque monkey visual area V1," *Investig. Ophthalmol. Vis. Sci. ARVO Abst.*, Vol. 34, p. 1173, 1993.

45. T. Yoshioka, J. B. Levitt and J. S. Lund, "Independence and merger of thalamocortical channels within macaque monkey primary visual cortex: Anatomy of interlaminar projections," in preparation.

46. S. Zeki and S. Shipp, "Modular connections between areas V2 and V4 of macaque monkey visual cortex," *Eur. J. Neurosci.*, Vol. 1, pp. 494-506, 1989.

# The large scale organization of the primate cortical visual system

Malcolm P. Young

University Laboratory of Physiology  
Parks Road  
Oxford OX1 3PT

## ABSTRACT

The primate cortical visual system is composed of many structurally and functionally distinct areas or processing compartments<sup>1,2,3,4,5</sup>, each of which receives on average about ten afferent inputs from other cortical areas and sends about the same number of output projections<sup>1</sup>. The visual cortex is thus served by a very large number of cortico-cortical connections, so that the areas and their interconnections form a network of remarkable complexity. The gross organization of this cortical processing system hence represents a formidable topological problem: while the spatial position of the areas in the brain is becoming fairly well established, the gross 'processing architecture', which is defined by the connections, is much less well understood. The problem arises because there are too many connections to sustain unaided intuitions about the organization of the system made on the basis of examining the primary connection data. Analysis of the connection data that shows the connectional organization of the visual system is required.

I have applied optimization analysis to connectional data on the cortical visual system to address this topological problem<sup>6</sup>. This approach gives both qualitative and quantitative insight into the connectional topology of the primate cortical visual system<sup>6</sup>, and provides new evidence supporting suggestions that the system is divided into a dorsal 'stream' and a ventral 'stream' with limited cross-talk<sup>7</sup>, that these two streams reconverge in the region of the principal sulcus (area 46) and in the superior temporal polysensory areas<sup>8,9</sup>, that the system is hierarchically organized<sup>1</sup>, and that the majority of the connections are from 'nearest-neighbour' and 'next-door-but-one' areas. The robustness of the results is shown by reanalyzing the connection data after various manipulations that simulate gross changes to the neuroanatomical database.

## 1. INTRODUCTION

Three experimental approaches have been important sources of information about the organization of the primate visual system. First, testing the behavioural effects of selective brain lesions has demonstrated aspects of the causal role of many brain areas in visual information processing. Second, neurophysiological analysis of the response properties of cells in different areas has shown the distribution of cells' preferences for particular visual features. Third, powerful anatomical techniques, mainly involving the injection of actively transported tracers, have revealed the visual areas to be connected by hundreds of ipsi- and contra-lateral cortico-cortical connections, and by an intricate subcortical network.

The latter kind of data is important because the connections define the pathways in which information may or may not flow. Representing the connectional organization of the visual system is particularly important because the nature of the input to a brain structure, and the effect of its output, depend in part on the 'place' of the area in the system of connections. Hence, there can be little possibility that the functions of different visual structures, the functions of interactions between them, or indeed the function of the processing system as a whole, can be properly understood without first characterizing the 'wiring' pattern. Despite the importance of connectional data in understanding the organization of the visual system, however, many interpretations of the data have taken an informal, casual and speculative form, often without any supporting analysis, in contrast to the convention in interpreting neurophysiological data and data from lesion studies.

One important exception to this speculative approach has been the application of hierarchical analysis to connectional data<sup>1,10,11</sup>. This analysis considers the cortical laminae in which connections originate and terminate. By assuming that terminations in cell-rich layers are 'ascending', and that terminations in cell-sparse layers are 'descending' it is possible to arrange the visual cortical areas into a largely consistent unidimensional hierarchy. This analysis indicates both that the cortical visual system is hierarchically organized, and the probable direction of the flow of signals in the system. It has some limitations, being dependent on detailed data on the laminar origin and termination of connections that are not available for many connections, less applicable to structures that do not have clear laminar organization, and not giving any insight into organizational features that are not hierarchical. The latter limitation means that this type of analysis does not speak to the issue, for example, of whether the visual system is divided into discriminable processing streams, as has been suggested by lesion studies and by neurophysiology<sup>7,12</sup>. The left-to-right positions of areas in the familiar hierarchical diagram of Felleman and Van Essen<sup>1</sup> could be shuffled at random without doing violence to the analytical rules and the data that constrain the diagram.

Because hierarchy may not be the only organizational feature of the visual system, and because the "identification" of other organizational features without any objective analysis is rather eccentric from the point of view of quantitative biology, I have described a further analytical approach to connectional data. This approach uses optimization to produce multidimensional representations of the organization of a brain system that can respect almost any connection pattern<sup>6,13,14</sup>. It can use widely available neuroanatomical data, but gives direct insight into the likely direction of flow of signals only where there are non-reciprocal connections<sup>14</sup>. The results of this newer approach and those from hierarchical analysis complement one-another.

## 2. OPTIMIZATION ANALYSIS OF VISUAL CORTICAL CONNECTIVITY

The optimization analysis begins by identifying the brain areas of interest, and proceeds by examining a matrix of connections between these areas. The values in a connection matrix are 'proximities' that define spatial relations between points representing the brain structures in a space. The defined spatial relations can be perfectly reflected in the configuration of points in this space, so that connected points are close together and unconnected ones far apart, only when the space has a large number of dimensions. The connectional organization can be made understandable by reducing the dimensionality of the space to three or fewer dimensions, while preserving as much as possible of the proximities between the points of the configuration. The low-dimensional configuration of points produced by the analysis optimally fits the connection matrix so that the proximities of the points of the structure are as close as possible to the rank order of the 'proximities' of areas in the connection matrix. This dimensional reduction is brought about by nonmetric multidimensional scaling (MDS)<sup>15,16,17</sup>.

The visual cortex can be divided into different areas according to several different parcellation schemes (e.g. refs 1, 4, 5). I used the most recent parcellation of the cortex into areas<sup>1</sup>, and examined a matrix of connections between these areas of the macaque cortical visual system. This matrix is set out in Table 1. The connection matrix was analysed by MDS, yielding a configuration of points that correspond to the cortical areas. The configuration optimally fits the matrix so that the length of known connections is at a minimum given the co-constraint that the length of presently unreported connections is at a maximum. Hence points representing areas which have very similar afferent and efferent cortico-cortical connections are close together, while points representing areas which have very different patterns of connections are far apart. The details of the analysis are indicated in the Figure 1 legend. Figure 1 shows the best-fit structure for the connectivity matrix in Table 1. Points representing the areas of the cortical visual system are shown connected by the projections to and from the corresponding areas.

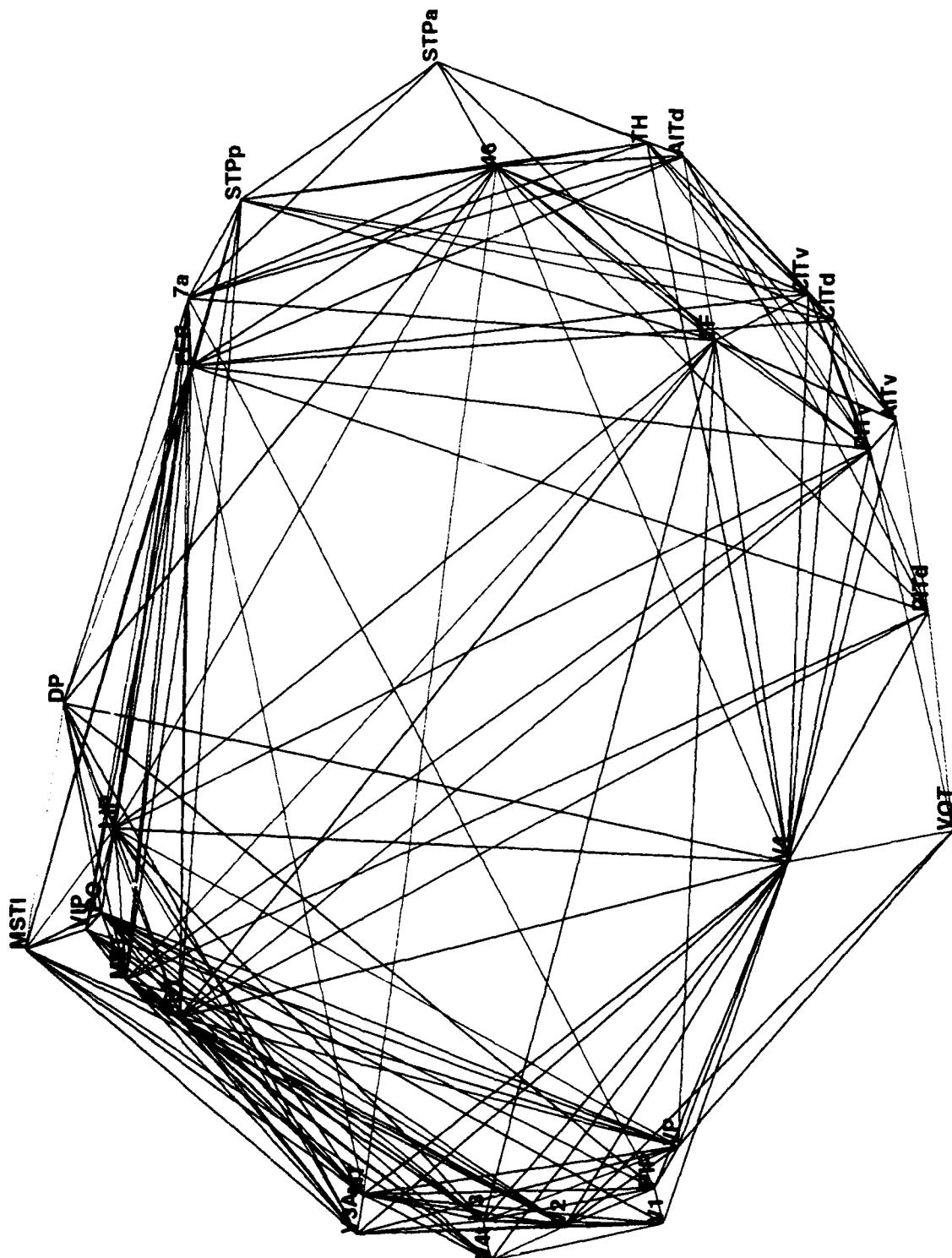
**Table 1.**

	V1	V2	V3	VP	V3A	V4	VOT	V4L	MT	MSTD	MSTI	FST	PITd	PITv	CITd	CITv	AITd	AITv	STPp	STPa	TF	TH	PO	PIP	LIP	VIP	DP	7a	FEF	46	
V1	0	1	1	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0		
V2	1	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0		
V3	1	1	0	0	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0		
VP	0	1	1	0	1	1	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0		
V3A	1	1	1	1	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0		
V4	1	1	1	1	1	0	0	1	0	0	1	1	1	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1	0		
VOT	0	1	0	1	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
V4L	1	0	1	0	0	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0		
MT	1	1	1	1	1	1	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1		
MSTD	0	1	1	1	1	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	1	1	1	1	1	0	0	0		
MSTI	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	0		
FST	0	0	1	0	1	1	0	1	1	0	0	0	0	0	1	0	1	0	0	0	1	1	0	1	1	0	1	0	0		
PITd	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	1		
PITv	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	1	1		
CITd	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	1		
CITv	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	1	1	1		
AITd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	1	1	1		
AITv	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0		
STPp	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	1	1	1	1	1	1		
STPa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	1	0	1	0	1		
TF	0	0	1	1	0	1	0	0	0	1	0	1	0	1	1	1	1	1	1	0	0	0	0	0	1	0	1	0	1		
TH	0	0	0	0	0	1	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	1	0	0	0	0		
PO	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0		
PIP	1	0	1	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0		
LIP	0	0	1	1	1	1	0	0	1	1	1	0	0	0	0	0	0	1	0	1	0	0	1	1	1	1	1	1	1		
VIP	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	0		
DP	0	0	0	0	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	1	0	1		
7a	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	1	1	1	1	0	1	0	1	0	1	1	1		
FEF	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	0	1	1	1	0	1	1		
46	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	1	1	0	0	0	1	1	1	1	1	0		

**Table 1:** Matrix of connections between areas of the macaque visual cortex. The cortical parcellation and connections are exactly as ref. 1 with the exception of MIP and MDP, which have been excluded. Connections coded as '1' are reported to exist and those coded '0' are either projections which have been explicitly tested for and found absent or connections which are not presently known. No information about the spatial position of the areas in the brain, the laminar patterns of the connections, the continuity or patchiness of the distribution of cells giving rise to a projection or about the relative density of projections is represented in the matrix. The information represented concerns only the existence of a connection between two areas, and is therefore the coarsest and most reliable that can be extracted from neuroanatomical studies. Nonetheless, some entries in the matrix can be expected to change as knowledge of the cortical areas and their connections is refined.

**Figure 1:** The topological organization of the macaque cortical visual system. Reciprocal connections are coloured red, one-way projections going from left to right are coloured blue and one-way projections going from right to left are green. A total of 301 connections is represented, of which 62 are one-way. This non-arbitrary structure is a best-fit representation in 2 dimensions of the connectional topology of this system, in which the positions of areas are specified by their positions being ones which minimize the distance between connected areas and maximize the distance between areas which are not connected. The analysis represents in a spatial framework the organizational structure of the network of cortico-cortical connections between elements of the visual cortex. In detail, the structure was derived by submitting the proximity matrix in Table 1 to non-metric multidimensional scaling<sup>15,16,17</sup>, using ALSCAL<sup>16</sup>. Solutions with the level of measurement specified as nominal and ordinal were derived, to assess whether a least-squares categorical transformation was required<sup>16</sup>, but there was no perceptible difference between them. Ordinal solutions in 1 to 5 dimensions were derived so that solutions with different dimensionality could be compared in a 'scree' test. This test showed diminishing returns in numbers of dimensions greater than 2. This structure was accordingly derived with an ordinal level of measurement in 2 dimensions, and the configuration of points (60 parameters) accounted for 40% of the variability in Table 1 (435 parameters). The results of the 'scree' test notwithstanding, some connections exist between areas which are widely separated in the structure, suggesting that these areas are topologically close in higher dimensions.

Figure 1.



Several features of the connectional organization of the system are immediately apparent. The two dimensions of the structure approximately correspond to the posterior-anterior (left to right in Figure 1) and to the dorsal-ventral (top to bottom) spatial distribution of the areas in the brain. For example, areas of the posterior parietal cortex and of the caudal superior temporal sulcus appear in the top part of the diagram, while areas of the inferotemporal cortex are located in the lower part. Because no information regarding the spatial position of the areas entered the analysis, only information concerning the areas to which each area is connected, this feature suggests that the spatial position of an area in the brain is a good predictor of the areas to which the area is likely to be connected, and that nearby areas tend to innervate one another (see below).

Beginning at the far left of Figure 1, where primary visual cortex (V1) is located, visual signals pass to a cluster of prestriate areas. This 'prestrial group' consists of areas V2, V3, VP, V4t, V3A, MT, and, perhaps surprisingly, area PIP. Areas V3A and MT are topologically less peripheral than other members of this group, and MT is only distinguished topologically by its (sparse) one-way projections to frontal cortex area 46 and to the frontal eye fields (FEF). Every area of the 'prestrial group' sends output connections to a further cluster of areas comprising FST, MSTd, MSTl, VIP, PO, LIP and DP. The projections from the 'prestrial group' to this group appear highly redundant, which might account for the fact that partial damage to these prestriate areas does not seriously disrupt spatial vision<sup>7</sup>. Signals from the 'MST/posterior parietal complex' then pass to the FEF, area 7a, the posterior part of the superior temporal polysensory area (STPp), and eventually to area 46 and the anterior STP (STPa).

Moving downwards from V1, signals are relayed via V4 and VOT into the inferotemporal (IT) cortex. As V4 is the principal gateway for signals entering IT, it would seem unlikely on topological grounds that V4 is involved only in colour vision (see refs 3,18,19). The IT cortex appears to be hierarchically organized, in the sense that more anterior stations are topologically further from the sensory periphery, and is associated with parahippocampal areas TF and TH. The topologically 'higher' areas of IT, where some cells respond with high specificity for particular visual patterns<sup>20-24</sup> project to area 46 and to STPa.

Connections between the dorsal and ventral streams are much less dense than those within each stream, and opportunities for cross-talk do not exist at every station. Both streams, however, project selectively to area 46 and to STPa. Area 46, for example, receives signals that presumably concern what an object is (from IT), where it is (area 7a, LIP), its movement in visual space (MT, MSTd, MSTl), its colour (V4) and its relation to movements of the eyes (FEF).

### 3. ROBUSTNESS OF THE RESULTS

The above analysis of the organization of the visual cortex only provides a compelling result if it is robust. The solution must be robust against two things. First, it must be robust against the differences in density or strength of the different projections: it is possible that when data relating to strong, moderate or weak connection densities are included, the solution will be markedly different. In fact, in no analysis of any sensory system of either the cat or monkey<sup>30</sup>, has the inclusion of this information given rise to a solution that explains less than 90% of the corresponding "binary" solution. Even if the data are treated as being metric (which they are not) and relatively large differences in density are introduced, the solutions remain similar, probably for two reasons. Connectivity is sparse, and so even a weak connection is a rare attractive constraint, and structures that are topologically close (i.e. those that have a very similar pattern of connectivity) tend to exchange strong connections.

Second, the solution must be robust against changes in status of some of the possible connections that have not so far been reported on: some of these connections will be found to exist. In fact, the most violent perturbation of the connection data, in which all unreported connections are assumed to exist, an assumption that would turn cortical neuroanatomy on its head, results in a solution that is 76% similar to that in Figure 1<sup>6</sup>. The solutions are constrained to be similar because a sufficiently large number of connections have been looked for and reported absent<sup>1</sup>. Poorly studied areas, like VOT and V4t, have their positions shifted by a large number of hypothetical new connections, but the organizational features of the solution are very similar. Thus, the grossest perturbation of the data does not disturb the conclusions, and they are unlikely to be overturned completely by growth in our information about visual cortical connectivity. Naturally the solutions will evolve, as the hierarchical diagrams have done, but the results from this type of analysis presently appear to be robust.

#### 4. QUANTITATIVE COMPARISON WITH OTHER RESULTS

I used the mathematical tractability of the structure in Figure 1 to investigate quantitatively whether the topological organization of the visual cortex reflects the dichotomy, reconvergence, hierarchy and border relations suggested by qualitative inspection, and by results from hierarchical analysis and lesion studies. This was accomplished by a regression-like procedure, Procrustes rotation<sup>25,26</sup>, which compared artificial model configurations that numerically embody each of the proposed organizational features, against the structure in Figure 1. This procedure finds the optimal reflection, rotation and scaling of each organizational model with the structure in Figure 1, and at the optimal comparison, yields a variance-explained statistic which reflects the goodness-of-fit between the two compared models. The statistical rarity of each comparison was assessed by an approximate randomization test<sup>27</sup>, which repeated the PROCRUSTES rotation with the organizational model shuffled randomly on each of 600 iterations. The number of times that the variance-explained statistic was exceeded during these random iterations was divided by the number of iterations to yield a probability that a correspondence as good as the particular comparison could have come about by chance.

Possible neighbourhood wiring rules embedded in the organization of the visual system were investigated by constructing matrices analogous to Table 1. The first matrix was derived by scoring hypothetical connections between areas which share a common border as '1' and all other possible connections as '0': the "nearest-neighbour" model (all connections were assumed to be reciprocal). This nearest-neighbour wiring matrix accounted for 61 out of 301 connections in Table 1 (27%). A second matrix was derived by scoring possible connections as a '1' if the areas share a common border or if they are separated by only one intervening area which abuts both areas: the "nearest-neighbour or next-door-but-one" model. Of the 301 connections in Table 1, 169 (55%) were connections between nearest-neighbour or next-door-but-one areas. Both these matrices were submitted to the same procedure as derived the structure in Figure 1, and the resulting configurations were compared with Figure 1 by Procrustes rotation.

The hierarchical ladder derived from the laminar origin and termination patterns of projections<sup>1</sup> was used to construct a unidimensional hierarchical model, by associating an integer value with each area according to its height above V1 in the Felleman and Van Essen scheme. This was the "hierarchical" model. To model the dichotomization of the system into a dorsal and ventral stream, and the subsequent reconvergence of these two streams, I assigned each area to one of three categories. Posterior parietal areas, caudal superior temporal areas and areas associated with eye movement (e.g. areas that would be thought part of the "dorsal stream" of ref 7) were assigned a '3'. 'Shared' areas, such as V1 and V2 (representing the 'shared' origins of the streams in occipital cortex), and STPa and area 46 (representing the reconvergence of the streams), were assigned a '2'. Areas of Ungerleider and Mishkin's ventral stream<sup>7</sup> were assigned a '1'. This was the "two streams and reconvergence" model. Finally, I derived a "combined hierarchical, two streams and reconvergence" model by making a 2-dimensional configuration in which the hierarchical model was dimension 1 and the two streams model was dimension 2.

**Table 2.**

Model	r <sup>2</sup>	probability
Nearest-neighbour	0.29	p < 0.0017
Nearest-neighbour or next-door-but-one	0.32	p < 0.0017
Hierarchical	0.30	p < 0.0017
Two streams and reconvergence	0.46	p < 0.0017
Combined hierarchical, two streams and reconvergence	0.72	p < 0.0017

Table 2 shows the results of the quantitative comparison of these models with the structure in Figure 1. All five models were related to the structure derived for the real cortical visual system at a level which would not be expected by chance (less than 1 in 600 probability). The two models that represented border relations between areas explained about the same amount of variability (30%) as the hierarchical model. The model that represented dichotomization and reconvergence, however, explained almost half the variability, while the combined model accounted for almost three quarters of the variability in Figure 1.

### 5. CONCLUSIONS

These results suggest that, despite the enormous complexity of the cortical visual system, at this gross level it may be organized according to four principles. (i) It is dichotomized into two streams, (ii) both streams are hierarchies, (iii) the streams reconverge in area 46 and STPa, and (iv) neighbouring areas tend to innervate one-another.

The finding that border relations may be present in the wiring pattern of the system might reflect the evolutionary advantage of keeping wiring to a minimum<sup>28,29</sup>, although the redundancy of connections from the prestriate areas to the areas of the caudal superior temporal sulcus and the posterior parietal cortex suggests that wiring economy is not a strong constraint. Alternatively, the tendency of areas to innervate their neighbours may reflect a parsimonious developmental process.

Two of these organizational features derived from topological analysis of the patterns of connections, the "two streams" and "hierarchical" features, corroborate organizational principles derived from different information sources, such as lesion studies<sup>7</sup> and from the laminar termination patterns of cortico-cortical projections<sup>1</sup>, respectively. In these cases, disputation that the visual system is hierarchical or that it is divided into two discriminable subsystems should presumably explain why analyses of completely different data by completely different methods should come to such similar conclusions.

### 6. ACKNOWLEDGEMENTS

Supported by a Royal Society University Research Fellowship, the MRC, the Wellcome Trust and the Oxford McDonnell-Pew Centre for Cognitive Neuroscience. I am grateful to C. Blakemore.

## 7. REFERENCES

1. Felleman D.J. and Van Essen D.C. "Distributed hierarchical processing in the primate cerebral cortex," *Cerebral Cortex* 1:1-47 (1991)
2. Peters A. and Jones E.G. *Cerebral Cortex* (Plenum, New York and London, 1985)
3. Zeki S. and Shipp S. "The functional logic of cortical connections," *Nature* 335:311-317 (1988)
4. Seltzer B. and Pandya D.N. "Afferent cortical connections and architectonics of the superior temporal sulcus and surrounding cortex in the rhesus monkey," *Brain Res.* 149:1-24 (1978)
5. Boussaoud D., Ungerleider L.G. and Desimone R. "Pathways for motion analysis: cortical connections of the medial superior temporal and fundus of the superior temporal visual areas in the macaque," *J. Comp. Neurol.* 296: 462-495 (1990)
6. Young M.P. "Objective analysis of the large scale organization of the primate cortical visual system," *Nature* 358:152-154 (1992)
7. Ungerleider L.G. and Mishkin M. "Two cortical visual systems," In: Ingle D.G., Goodale M.A. and Mansfield R.J.Q. (Eds.) *Analysis of Visual Behavior*. 549-586 (MIT, Cambridge MA, 1982)
8. Baizer J.S., Ungerleider L.G. and Desimone R. "Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques," *J. Neurosci.* 11:168-190 (1991)
9. Morel A. and Bullier J. "Anatomical segregation of 2 cortical visual pathways in the macaque monkey," *Vis. Neurosci* 4: 555-578 (1990)
10. Rockland K.S. and Pandya D.N. "Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey," *Brain Res.* 179:3-20 (1979)
11. Maunsell J.H. and Van Essen D.C. "The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey," *J Neurosci* 3:2563-2586 (1983)
12. Merigan W.H. and Maunsell J.H.R. "How parallel are the primate visual pathways?" *Annu. Rev. Neurosci.* 16:369-402 (1993)
13. Scannell J.W. and Young M.P. "The connectional organization of neural systems in the cat cerebral cortex," *Current Biology* 3:191-200 (1993)
14. Young M.P. "The connectional organization of neural systems in the primate cerebral cortex," *Proc. Royal Society: Biol. Studies* 252:13-18 (1993)
15. Young F.W. *Multidimensional Scaling: History, Theory and Applications*. Hamer R.M. (Series Ed.) (Hillside, N.J., 1987)
16. Young F.W. and Harris D.F. "Multidimensional scaling," In: Norusis M.J. (Ed.), *SPSS Base System User's Guide*, (SPSS inc., Chicago, 1990)
17. Shepard R.N. "Multidimensional scaling with an unknown distance function" *Psychometrika* 27:125-140 (1962)

18. De Yoe E.A. and Van Essen D.C. "Concurrent processing streams in monkey visual cortex," *Trends Neurosci.* 11:219-226 (1988)
19. Schiller P.H. and Lee K. "The role of the primate extrastriate area V4 in vision," *Science* 251:1251-1253 (1991)
20. Miyashita Y. "Neuronal correlate of visual associative long-term memory in the primate temporal cortex," *Nature* 335:817-820 (1988)
21. Young M.P. and Yamane S. "An analysis at the population level of the processing of faces in the inferotemporal cortex," In: Ono T., Squire L.R., Fukuda M. and Perrett D.I. (Eds.) *Brain Mechanisms of Perception and Memory: from Neuron to Behavior.* (OUP, in press)
22. Young M.P. and Yamane S. "Sparse population coding of faces in the inferotemporal cortex," *Science* 256:1327-1331 (1992)
23. Tanaka K., Saito H., Fukada Y. and Moriya M. "Coding visual images of objects in the inferotemporal cortex of the macaque monkey," *J. Neurophysiol.* 66: 170-189 (1991)
24. Gross C.G., Rocha-Miranda C.E. and Bender D.B. "Visual properties of neurons in the inferotemporal cortex of the macaque," *J. Neurophysiol.* 35:96-111 (1972)
25. Schoneman P. and Carroll R.M. "Fitting one matrix to another under choice of a similarity transformation and a rigid motion," *Psychometrika* 31:1-10 (1970)
26. Gower J.C. "Statistical methods of comparing different multivariate analyses of the same data," In: *Mathematics in the Archeological and Historical Sciences.* 138-149 (Edinburgh, 1971)
27. Edgington E.S. *Randomization Tests.* (Marcel Dekker, N.Y., 1980)
28. Mitchison G. "Neuronal branching patterns and the economy of cortical wiring," *Proc. Royal Soc. Biol. Studies* 245:151-158 (1991)
29. Cowey A. "Cortical maps and visual perception: the Grindley memorial lecture," *Q.J.Exp.Psychol.* 31:1-17 (1979)
30. Scannell J.W., Young M.P. and Blakemore C. "Optimization analysis of the connections between areas of the cat cerebral cortex," *Soc. Neurosci. Abs.* 18:313.10 (1992)

# **Dynamic object-based 3-D scene analysis using multiple cues**

**Teri B. Lawton**

**Nano Tech Services, 3700 Peninsula Road,  
Channel Islands Harbor, California 93035**

## **ABSTRACT**

A Computational Visual System (CVS) has been developed that segments objects in natural scenes using algorithms and filtering elements similar to those used by people. The filtering elements of the CVS are based on neural networks elucidated by physiological and anatomical studies. The algorithms of the CVS are based on data from psychophysical studies. This CVS classifies different types of patterns, based on object shape, texture, position in the visual field, and amount of motion parallax in subsequent scenes, without any *a priori* models. When analyzing 3 Dimensional (3-D) scenes, both psychophysical and physiological evidence indicates that people construct an object-based perception, one that is event-driven. The object-based representation being modeled focuses on the object formation found in the dorsal cortical pathway, used to locate an object in 3-D space. Therefore, the interaction between the eye-head movement system and the pattern recognition system is modeled. Both global scene attributes used to reveal objects masked by shadows and improve object segmentation, and local object attributes defined by the boundary of contrast differences between an object and its background are modeled. The importance of using paired odd- and even- symmetric detectors to form the boundary and analyze the texture of an object is emphasized. This information is used to construct a viewer-centered object-based map of the scene that is based on multiple object attributes. Algorithms that incorporate the relative weighting of the different object attributes being used to discriminate objects are used to instantiate computational networks that incorporate both competitive and cooperative networks. The current CVS enables one to: 1) test the effectiveness of different types of global and local filtering for improving object segmentation by visual inspection of the filtered scenes and object data in multiple windows, and 2) generate objects that have been segmented by the CVS to be used as stimuli in pattern discrimination experiments in natural scenes, a task requiring multiple cortical areas. This CVS can be used to improve: 1) understanding the algorithms used to: a) locate an object in 3-D space, and b) construct an elaborated wide field view of objects in natural scenes, by normal observers and those with cognitive deficits, and 2) automated pattern recognition systems useful for aiding navigation of partially sighted people and robotic vehicles.

## **1. INTRODUCTION**

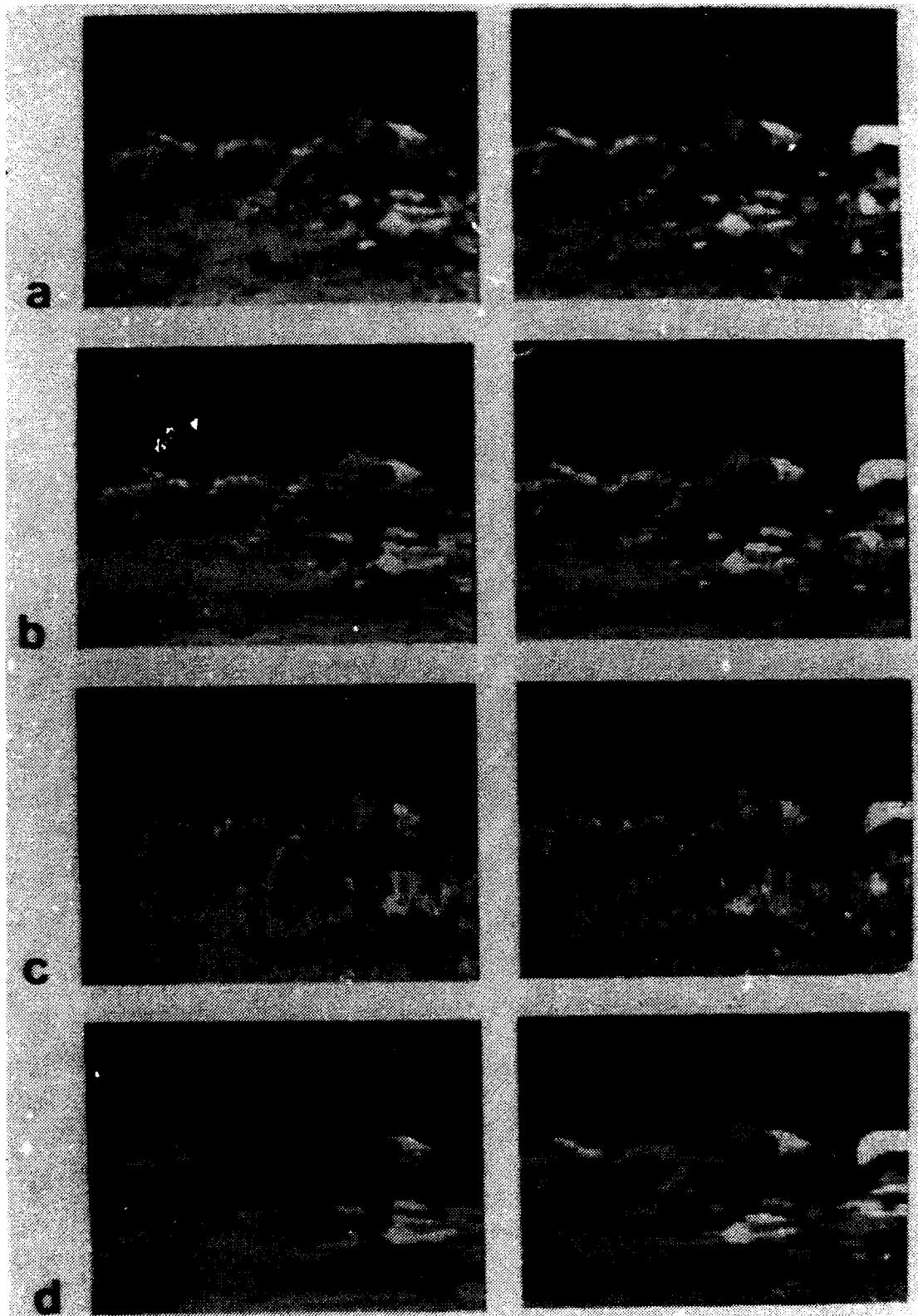
Natural scenes provide richly textured backgrounds that serve to camouflage certain objects in the scene. Currently, there is no software or hardware platform that is available to test an observer's ability to discriminate objects using parametric approaches when objects are embedded in natural scenes. In addition, the robustness of multiattribute models to predict object discrimination in natural scenes without any *a priori* object models cannot be tested using any available computer systems. Therefore, research in this area is scarce or nonexistent. Object segmentation in natural scenes is needed to: 1) construct stimuli that can be used for conducting psychophysical and physiological studies of pattern discrimination in natural scenes, enabling us to better understand how multiple object attributes are combined to construct a 3-D perception, 2) use object and depth discrimination thresholds to provide normative data for early detection of cognitive disorders, and 3) develop automated pattern recognition systems that can help guide robotic vehicles and partially sighted people. Science and technology have made significant advances in the last 20 years to enable: 1) More in depth understanding of biological neural networks, and 2) High speed, inexpensive digital hardware readily available to implement dynamic computational neural networks in the laboratory. Therefore, the capability exists to create robust real-time automated object discrimination and pattern recognition systems.

I have developed a 3-D Computational Visual System (CVS) that uses multiple object attributes and scene cues to construct 3-D object maps. This CVS uses algorithms and filtering elements similar to those used by people.<sup>33,34</sup> This 3-D CVS classifies different types of objects in natural outdoor scenes, based on object shape, texture, position in the visual field, and amount of motion parallax in subsequent scenes. This is done without any *a priori* object models. The algorithm development is based on three lines of research that have evolved to generate efficient algorithms: primate neurobiology, visual psychophysics, and computational vision. First, primate neurobiology has shown that the visual cortex is served by a large number of cortico-cortical connections, so that the different visual areas or processing compartments and their interconnections form a network of remarkable complexity.<sup>14,63</sup> Neurophysiological analyses of the response properties of cells in different visual areas have shown they respond optimally to different object attributes. Second, visual psychophysics studies have shown that several object attributes are important for object segmentation, including the use of: 1) paired odd- and even-symmetric filters to extract: a) the position of contrast boundaries,<sup>28</sup> and b) the gray scale texture of an object that is demarcated by the boundary, 2) a textured background frame of reference (both the value and spacing of spatial frequencies are analyzed) to judge the direction of movement,<sup>28</sup> and 3) a gradient analysis when movement occurs within the background frame of reference.<sup>32</sup> It is likely that these object attributes are analyzed predominantly by magnocellular pathways.<sup>32</sup> Third, inexpensive fast hardware having pipelined graphics, Xwindows and C software, and structured, modular, interactive program design using inline computations makes it possible to develop an event-based CVS that first, operates at close to real-time with no special image processing hardware boards being needed, and second, generates multiple windows for viewing the effects of different object segmentation algorithms, viewing any details of the analysis desired, the results being available in separate windows on the screen.

The filtering components of the CVS at each level of processing are constrained by neurophysiological data, whereas the algorithms are constrained by psychophysical data, both of these components being implemented using efficient software code. Robust design principles that are based on those selected by Darwinian evolution are used. Therefore, elaborate preprogramming is not needed. This CVS implements adaptive, dynamic, event-based scanning of the scene to incorporate a rudimentary attentive component and improve the robustness and speed of constructing object maps by the CVS. This layered neural network architecture provides an efficient, fast means for data reduction to construct the 3-D topographic layout of terrain from 2-D images. As a result of using sensory fusion, noise reduction, event-based sampling, and learning, partial information can be used to construct a robust depth map. By combining the sensory cues and inferred object attributes along several sensory dimensions, the segmentation of the scene into objects at different depths is more fault tolerant than other CVSS.

When analyzing 3-D scenes, both psychophysical and physiological evidence indicates that people construct an object-based perception,<sup>18,49,57,60,64</sup> one that is event-driven, including both attentive and preattentive processing of visual information.<sup>58,59</sup> The object-based representation being modeled in the current CVS focuses on the object formation found in the dorsal cortical pathway, used to locate an object in 3-D space. Therefore, the interaction between the eye-head movement system and the pattern recognition system is modeled. The eye-head movement system that is modeled compensates for observer movement, so that only translational movement varies, enabling the same object to be identified easily in subsequent views of the same scene,<sup>33,34</sup> see Fig. 1. Since depth is easily computed from motion parallax that is based on translational movement, motion parallax is the primary cue used to extract depth by this CVS. Motion parallax assumes that objects moving faster are closer than objects moving more slowly in subsequent views of the same scene.

Noise reduction and global optimization are used to improve object segmentation. This type of optimization will enable the CVS to resolve objects when viewed on low resolution displays or when objects are obscured by scene parameters, such as dust, rain, snow, low clouds, or high glare. Taking



**Figure 1.** (a) Two Views Of Natural Scenes Compensated for Pitch, Roll, Heading, and Contrast, (b) Horizontal Segment Maps, (c) Vertical Segment Maps, and (d) Object Maps That Were Constructed From These Scenes.

into account different scene parameters is a capability not possible with current pixel-based analyses, yet a type of obscuration that occurs frequently in some environments. Both global scene attributes used to reveal an object masked by shadows and improve object segmentation, and local object attributes defined by the boundary of contrast differences between an object and its background are modeled. The brightness of overlapping horizontal and vertical image segments are matched recursively to construct each object. Paired odd symmetric detectors to form the boundary and even symmetric detectors to analyze the texture of an object are incorporated in the model. This information is used to construct a viewer-centered object map based on multiple object attributes. Depth based on motion parallax is easily computed once the difficult problem of object segmentation has been solved.

## **2. PREVIOUS APPROACHES**

Image segmentation is the process of partitioning a digital image into disjoint connected sets of pixels, each of which corresponds to an object or region (see Castleman<sup>8</sup> for a review). Image segmentation can be approached either as the process of assigning pixels to objects or of finding boundaries between objects. Gray level thresholding is a simple segmentation technique that only works for objects on uniform backgrounds. It is necessary to allow the threshold gray level to vary within the image to accommodate changes in the background gray level. Commonly used techniques for object segmentation, such as determining a local minimum in the gray level histogram, the maximum average boundary gradient, or inflection points when computing the area or perimeter based on equal gray level, are too simple to segment objects in natural scenes. Object segmentation has been implemented by tracking the boundaries in the gradient image or by thresholding the gradient image. Boundary tracking, however, is very sensitive to noise in the scene, and requires random access to all pixels, a format not easily available when raster images are input. Region growing techniques are useful for complex scenes and complex object definitions. The segmentation of an image may be stored by a membership map, by a boundary chain code, or by line segment encoding. Line segment encoding enables features such as the object's area, perimeter, texture, average gray level, to be built into the object extraction step, generating the most efficient object representation.

There have been three major approaches for scene analysis using digital image processing. The first of these approaches uses artificial intelligence techniques based on a world model.<sup>3</sup> The second major approach for automated pattern recognition uses exhaustive pixel-based object segmentation, using a sequence of small filters to find edges at various orientations.<sup>8,21,38</sup> The third major approach uses computational neural networks, implementing a massively parallel computational architecture, starting with a random configuration of elements and back propagation to correct the subsequent computational errors. These neural network models do not simulate adult biological networks, since these networks operate very slowly, taking between 2000-8000 patterns to train the computational elements,<sup>39</sup> before they are able to complete a simple problem such as determining the direction an object is moving when viewed through an aperture.

These approaches have many limitations. It is hard if not impossible to represent complex unfamiliar scenes. These approaches are based on the intrinsic properties of the scene, such as the reflectance of visible surfaces, the geometric distribution and organization of intensity changes, and the observer's viewpoint, and not on the multiple sensory dimensions analyzed by biological systems. Thus, other data sources, such as the amount and type of observer movements, scene illumination, region growing, shadows, obscuration, and the resulting feedback that is needed for effective adaptive thresholding and unsupervised learning to create an object-based instead of a pixel-based representation of the scene, are ignored. None of these CVSSs model the interaction between the eye-head movement and the pattern recognition system. Therefore, translational movement of an object in subsequent views of the same scene cannot be used to construct the depth map. Pixel-based cross-correlation schemes use exhaustive scanning, adjusting the interconnection weights of individual nodes in the computational architecture on each iteration (regularization), instead of using event-based analyses to construct an object-based representation. Multiplicative cross-correlation analyses, in

addition to being much less robust than gradient analyses, are also much less computationally efficient. Therefore, previous computational models are not able to analyze scenes at real-time frame rates.

### **3. NEURAL NETWORK ARCHITECTURE FOR DYNAMIC SCENE ANALYSIS**

Biological systems evolved to acquire visual information rapidly. Movement is used by the visual system to segment the scene into separate objects and to break camouflage.<sup>24,25</sup> The eye-head movement system and the pattern recognition system work together to facilitate object segmentation. The observer's ocular motor system can compensate for the pitch, heading, and roll of the eyes and head relative to the observers viewpoint, so that only the translational movement remains. Translational movement is used to compute motion parallax,<sup>20</sup> a powerful cue used by the pattern recognition system to localize the position of an object in 3-D space. Depth is easily computed once the translational movements of objects have been extracted.<sup>36</sup>

There are neural circuits that link multiple cortical areas to thalamic and lower brainstem structures. These circuits control lower levels of processing such as eye movements and higher levels of processing such as attention.<sup>1,16,51</sup> These neural circuits are regulated by visual input from higher levels of cortical processing,<sup>52</sup> including areas such as Medial Temporal (MT) and Medial Superior Temporal (MST) cortex<sup>40,47,62</sup> where more global characteristics of motion such as motion parallax are analyzed. These neural circuits serve as a gain-control system, adjusting the activity levels in the basic circuit and its side-loops to scale movements in time and space.<sup>16</sup> This gain control is implemented in the current CVS by adjusting both global and local filtering parameters.

Objects are segmented using direction selective localization of object boundaries. Discriminating the direction of movement is a task determined initially in the cortex by oriented paired even and odd symmetric simple cells.<sup>22,50,53</sup> Paired even and odd symmetric simple cells act like bandpass channels tuned to approximately a 1 to 1 1/2 octave band of spatial frequencies found using psychophysics<sup>4,5</sup> and physiological recordings.<sup>12,45</sup> The paired odd and even symmetric simple cells in the striate cortex can be modeled using Gabor filters (which correspond to a sine x Gaussian and a cosine x Gaussian).<sup>11,37,50</sup> These are the basic filtering components in the current CVS.

There are several advantages that result from using these paired filters. For linear systems, these paired even and odd symmetric filters: 1) optimize resolution, requiring the smallest number of filters for scene analysis, in terms of both spatial position and spatial frequency,<sup>26</sup> 2) provide an orthogonal encoding scheme that describes the output of simple cells in the visual cortex,<sup>26,37,50</sup> 3) maximize the signal-to-noise ratios given a fixed number of cells,<sup>15,56</sup> and 4) provide a mechanism for implementing a gradient analysis to localize the position of object boundaries.

Psychophysical studies of phase discrimination elucidate additional stimulus conditions that optimize direction selective localization of object boundaries. The fundamental frequency of the background grating is used as the frame of reference for discriminating the direction of movement.<sup>28,32</sup> As the fundamental frequency of the background is lowered, direction discrimination occurs at lower contrasts for a wider range of different frequency patterns. In addition, if the test frequency was a harmonic of the background's fundamental frequency, then direction discrimination occurred at half the contrast needed for similar test patterns that did not repeat an integral number of times within the background frame of reference. These spatial phase discrimination thresholds were unaffected by pattern jitter, different background contrasts, short durations, and short intervals between pattern presentations,<sup>32</sup> indicating that the direction discrimination mechanism is mediated predominantly by magnocellular pathways, as is proposed by others.<sup>22,41,61</sup>

Both direction discrimination<sup>32</sup> and velocity discrimination<sup>30,42</sup> between patterns seen in two time intervals that are moving across the same region of space do not change as a function of background contrast, verifying the predictions of a gradient model. A linear gradient model is less susceptible than a multiplicative cross-correlation model to scene noise. Therefore, an object can be tracked much more precisely and quickly using a gradient model. A gradient model to predict movement discrimination of objects seen against simple complex backgrounds<sup>32</sup> is incorporated in the current 3-D CVS. The gradient model takes advantage of finding that observers are optimally sensitive to 90 deg spatial phase differences,<sup>27</sup> and that both global and local contrast differences<sup>28,32</sup> are used to discriminate movement.

Each area in the visual cortex is specialized to extract different object attributes, such as object boundaries, texture gradients, binocular disparity, and motion parallax relative to a background frame of reference.<sup>13,23,41,61,65</sup> Substantial evidence exists that there are two different streams of information, a ventral or predominantly parvocellular stream with limited magnocellular input, and a dorsal or predominantly magnocellular stream,<sup>13,22,41,43,60,61</sup> that connect different areas in the visual cortex in a hierarchical manner,<sup>13,55,61,65</sup> these streams having limited crosstalk until they reconverge in the region of the principle sulcus (area 46) and in the superior temporal polysensory areas, e.g. STS.<sup>44,60</sup> The ventral system has been found to convey color and form information, whereas the dorsal system conveys motion and depth information.<sup>13,41,60</sup> More local object attributes such as boundaries defined by luminance contrast and direction of movement are analyzed at low levels in the hierarchy, e.g. in the striate cortex, and more global object attributes, such as motion parallax, at a higher level of analysis, such as in MT and MST cortex,<sup>61,62</sup> whereas an object-centered perception is determined at the highest level of analysis.<sup>18,49,57,64</sup>

There is both feedforward and feedback connections between different visual areas and within the same area, with neighbouring areas tending to innervate one another.<sup>63</sup> There are interactions between different object attributes when constructing a 3-D object map, for example between binocular disparity, texture, shape from shading, occlusion, and motion parallax, depending on their relative strengths.<sup>7,9,48</sup> Global and local analyses involving cooperative and competitive processes among the multiple sensory dimensions<sup>10,19,38,58,59</sup> are used by biological neural networks to construct a multidimensional object map of the scene. This analysis minimizes the effects of noise<sup>46</sup> by having information along several stimulus dimensions analyzed concurrently in different areas of the brain. Simultaneously analyzing several perceptual dimensions improves the accuracy of dynamic scene analysis. The current CVS aids in determining the algorithms people use to combine information from multiple cortical areas to locate an object in 3-D space, by providing benchmarking tools that enable the user to see how well different adjustments to the algorithm parameters improve object segmentation.

#### **4. ADVANTAGES OF CVS**

The algorithm development is based on three lines of research that have evolved to generate efficient algorithms: primate neurobiology, visual psychophysics, and computational vision. The advantages of segmenting natural scenes to construct a depth map using algorithms based on biological systems can be seen by examining Fig. 2. Instead of using discontinuities in image intensities, an object-based representation is used to construct 3-dimensional terrain maps. Instead of using circularly symmetric filters that can be off by a factor of two, oriented odd and even symmetric filters are used to localize the position of boundaries as accurately as people can. A linear gradient analysis, providing a more accurate and faster localization of boundaries than a cross-correlation analysis, is used to determine: 1) an object's boundaries, 2) the same object in subsequent scenes, and 3) its relative depth. Thus, the algorithms analyze both spatial gradients to detect object boundaries, and temporal gradients to detect movement in subsequent views of the same scene. Several different object attributes such as width, height, grayscale, texture, shadows, and motion parallax are used to

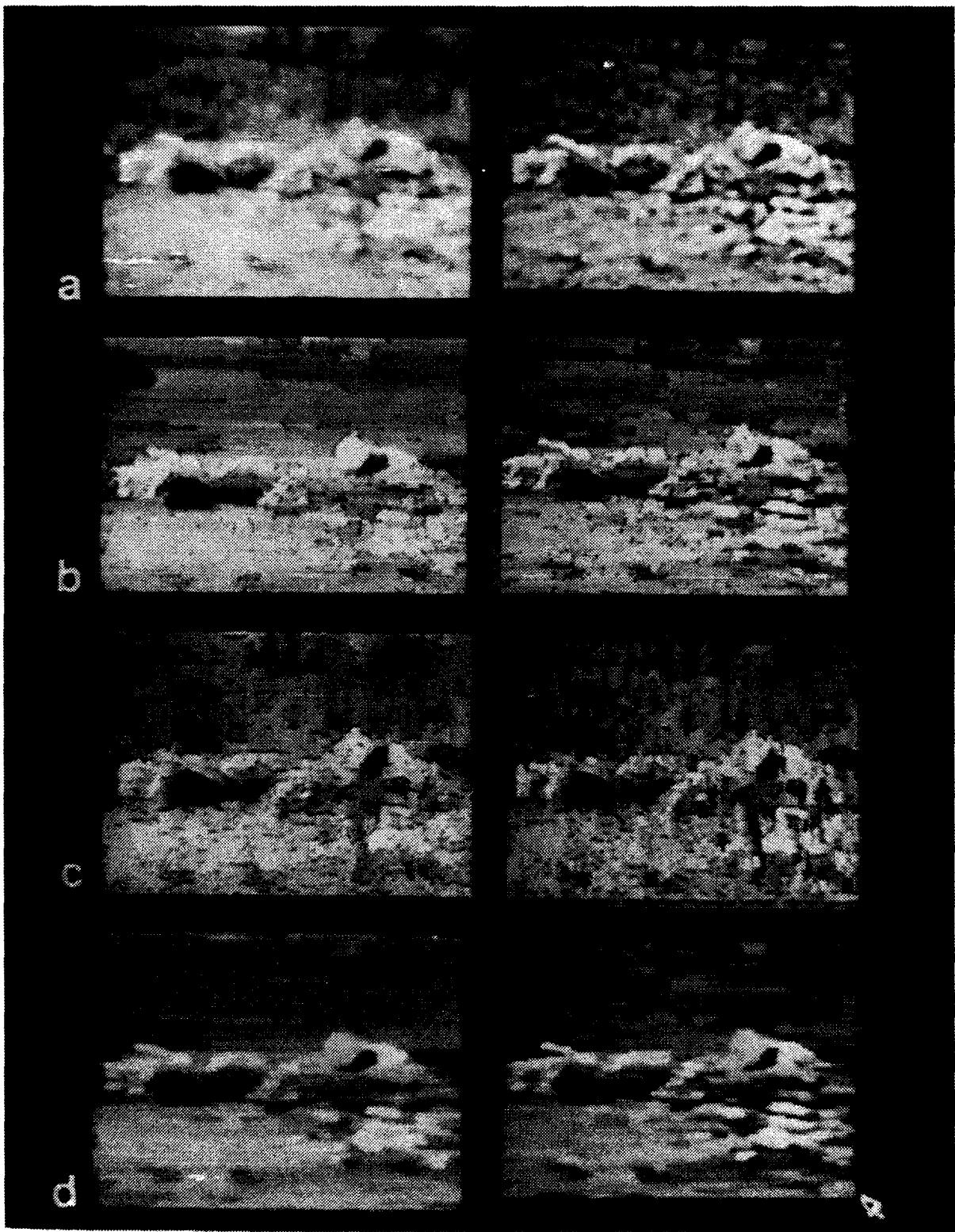
construct each object, compensating for the lack of precision along a single dimension (e.g. spatial phase or position) when matching the same object in subsequent scenes. Therefore, partial information along one dimension does not significantly degrade the object map construction. Since objects are segmented and matched in subsequent scenes based on the same pattern of activity from paired odd and even symmetric filters, object segmentation is done using scale invariant analyses. Region growing and event-based sampling based on the output of paired odd and even symmetric filters to construct an object-based terrain map, instead of relying on a pixel-based map, minimizes the effects of scene noise, providing a means for global optimization.

Type of Processing	Old Way	New Way
<b>Scene Representation</b>	<b>Pixel-Based</b>	<b>Object-Based</b>
<b>Filters</b>	<b>Circularly-Symmetric</b>	<b>Oriented Even- and Odd-Symmetric</b>
<b>Object Matching</b>	<b>Cross-Correlation</b>	<b>Gradient</b>
<b>Computations</b>	<b>Serial</b>	<b>Parallel and Serial</b>
<b>Sensory Modularity</b>	<b>Preattentive Vision Using Intrinsic Scene Properties</b>	<b>Both Preattentive and Attentive Vision Using Sensory Fusion that is Based on Biological Processes</b>
<b>Depth Extraction</b>	<b>Static Stereo</b>	<b>Dynamic Motion Parallax, Multiple Object Attributes (Effects of Shadows, Scene Noise, and Occlusion reduced)</b>
<b>Feedback</b>	<b>Limited, Not Dependent on Sensory Fusion</b>	<b>Controls Dynamic Working Range and Adaptive Thresholds (Based on Psychophysics)</b>
<b>Scanning</b>	<b>Exhaustive</b>	<b>Event-Based Subsampling Within Variable Windows of Attention</b>
<b>Learning</b>	<b>None, Supervised, or Bayesian</b>	<b>Unsupervised Event-Based Using Multiple Object Attributes</b>

**Figure 2. Comparison of Approaches for Computational Vision**

Initial benchmarking<sup>33,34</sup> found that high resolution object maps can be produced in 10-12 seconds, at least two to three orders of magnitude faster using the current CVS than with conventional pixel-based cross-correlation analyses using the same hardware. The thresholds for segmenting the scene into different objects using object-based analyses are designed to be adaptive, depending on the mean luminance in different windows of attention, whereas most other vision systems use fixed thresholds. Natural scenes, instead of synthetic images, can be used to determine the relative weighting of object attributes to locate an object in 3-D space. Using natural scenes will prevent the inaccurate weighting of cues for depth discrimination between, for example, binocular disparity and perspective, that is obtained when using synthetic images. The adaptive thresholding parameters used to filter both local object attributes (boundaries or texture gradients) and global scene attributes are easily changed.

The software is very modular and efficient, being optimized using C inline code, providing the tools needed to construct objects from lists of segments. The results of different object segmentation algorithms can be viewed on the screen in multiple windows, where the original image, filtered versions of the original image, the corresponding horizontal and vertical segment maps, and the

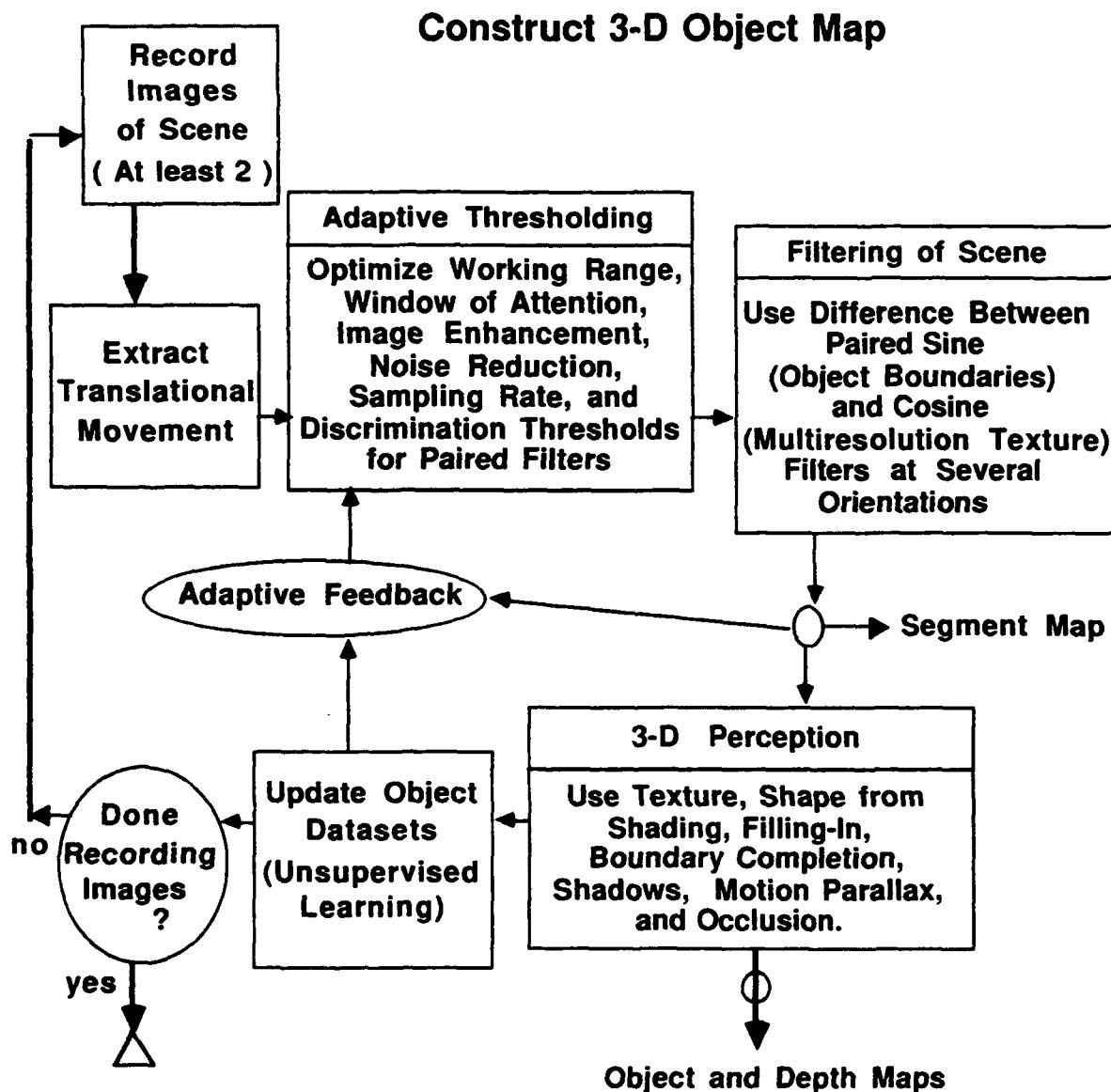


**Figure 3.** (a) Unenhanced (on left side) and Enhanced (on right side) Views Of Same Natural Scene Compensated for Pitch, Roll, Heading, and Contrast, (b) Horizontal Segment Maps, (c) Vertical Segment Maps, and (d) Object Maps That Were Constructed From These Scenes.

resultant object map, can be displayed, providing a very useful tool, e.g. see Figs. 1 and 3. The gray level of each pixel in the image window is printed in a data window when the mouse button is pushed, and the image is doubled in size or zoomed, at the location of the mouse in the image window, when a key on the keyboard is pushed. These windows are forked processes that exist independently of the program to construct the segment, object, and depth maps. Therefore, these windows can be saved and compared with subsequent analyses to improve the object map, e.g. see Fig. 6.

Object segmentation can also be used to provide stimuli for psychophysical and physiological studies of pattern discrimination when multiple cortical areas are used, for example, when constructing a 3-D perception of natural scenes. Since the algorithms being implemented will be based on psychophysical, physiological, and anatomical data of object and depth discrimination in natural scenes, then increasing the robustness of this CVS should help uncover how the mechanisms used by the visual system locate an object in 3-D space.

## 5. CURRENT CVS



**Figure 4. Computational Vision System to Construct 3-D Object Map of Scene.**

Currently, a dynamic object-based representation of a natural scene is constructed that uses the techniques of 1) boundary detection based on the output of oriented paired odd-symmetric (sine) and 2) luminance detection based on the output of even-symmetric (cosine) bandpass filters, see Fig. 4 for an overview of the CVS. To implement the filtering, each filter is centered on each pixel in the image, and the output of the filtering is measured using an odd number of cells or taps in the filtering kernel, for example, a 7x1 multiplier that is scaled to filter a 1 octave band of spatial frequencies, as in Figs. 1, 3, and 6b. The output of the odd-symmetric filter is always centered around the object's mean luminance, by its very nature. Therefore, using paired even and odd symmetric filters shifts the discrimination working range for object segmentation around the object's mean luminance, providing one type of automatic adaptive thresholding, and spurious results found when using zero-crossings (zero-crossing filters are unable to detect an edge whose luminance is restricted to be above the pattern's mean luminance)<sup>11</sup> cannot occur. The endpoints of each segment are determined by the location of the peak in the sine output from the filtered image. The endpoints of the segment are placed in the middle of the inflection point where the output of the sine filter is at a maximum. This boundary detector is consistent with psychophysical data, since Lawton<sup>32</sup> showed that the contrast needed to detect object boundaries, predicted using a gradient model consisting of the sum and difference of paired odd and even symmetric filters, could be expressed solely in terms of the contrast threshold of the odd symmetric detector. The average grayscale of each segment is determined by measuring the cosine filter times the grayscale of each pixel in the middle two-thirds of the segment. Whenever two adjacent segments are close together in luminance, then the segments are combined and the mean and sd of the segment's grayscale is updated. Each segment has a minimum size. Segments are coalesced together if either they are close together in grayscale, or if the length of the segment is below the size threshold.

The closed boundary contour of an object is defined by the path where the outputs of the sine filters are maximum, computed using horizontal and vertical orientations. An object is constructed, using a recursive cooperative analysis, by collecting all overlapping horizontal and vertical segments matching in grayscale. Gray-scale is matched based on the standard deviation (sd) of the pixels used to comprise all the segments in each object, and a texture threshold (tt), i.e.

if ( |gs(segment) - gs(object)| < sd(gs(object)) + tt ), then add segment to object. The sd of the smoothed grayscale values provides one measure of the object's texture, the constant tt provides a second measure that can take into account different types of scene illumination and scene noise. The texture threshold is used for filling-in or coalescing, so that regions of similar grayscale are grouped together. Objects that are too small in terms of width, height, density, or contrast using two different criterion levels, are combined with the adjacent object having the closest mean gray level. An object is assigned the average gray level of all included segments.

A high speed, 16 million instructions per second (mips), high resolution, 1160 x 900 pixels, Sun IPC SPARCstation using Xwindows X11-R5 implemented on a C platform, having 256 levels of gray, was used to implement the algorithms and the benchmarking tools that were used to test the robustness of the CVS. Stimulus scenes were constructed from a sequence of video images taken with a video camera that was attached to an automated gyroscope, inclinometer, and compass to enable compensating each scene for the pitch, heading, and roll of the camera.<sup>33,34</sup> Thus, only translational movement varies from one scene to the next. The natural scenes were filtered by maximizing the range of gray levels using a cumulative luminance histogram and bilinear interpolation,<sup>8</sup> e.g. see Figs. 1a, 3a, and 6a.

Image enhancement filters that boost the amplitude of low spatial frequencies more than intermediate and high spatial frequencies using a nonlinear Wiener filter,<sup>29,31,35</sup> were used to compensate for the low resolution video image. These filters revealed objects hidden beneath shadows, Fig. 5b, and also improved object segmentation, Figs. 3 and 6 (images on right side). Objects hidden beneath shadows are uncovered, since visual information in the shadows is concentrated at the low spatial frequencies. When the filtering parameters are carefully adjusted, Figs. 3d and 6b show that

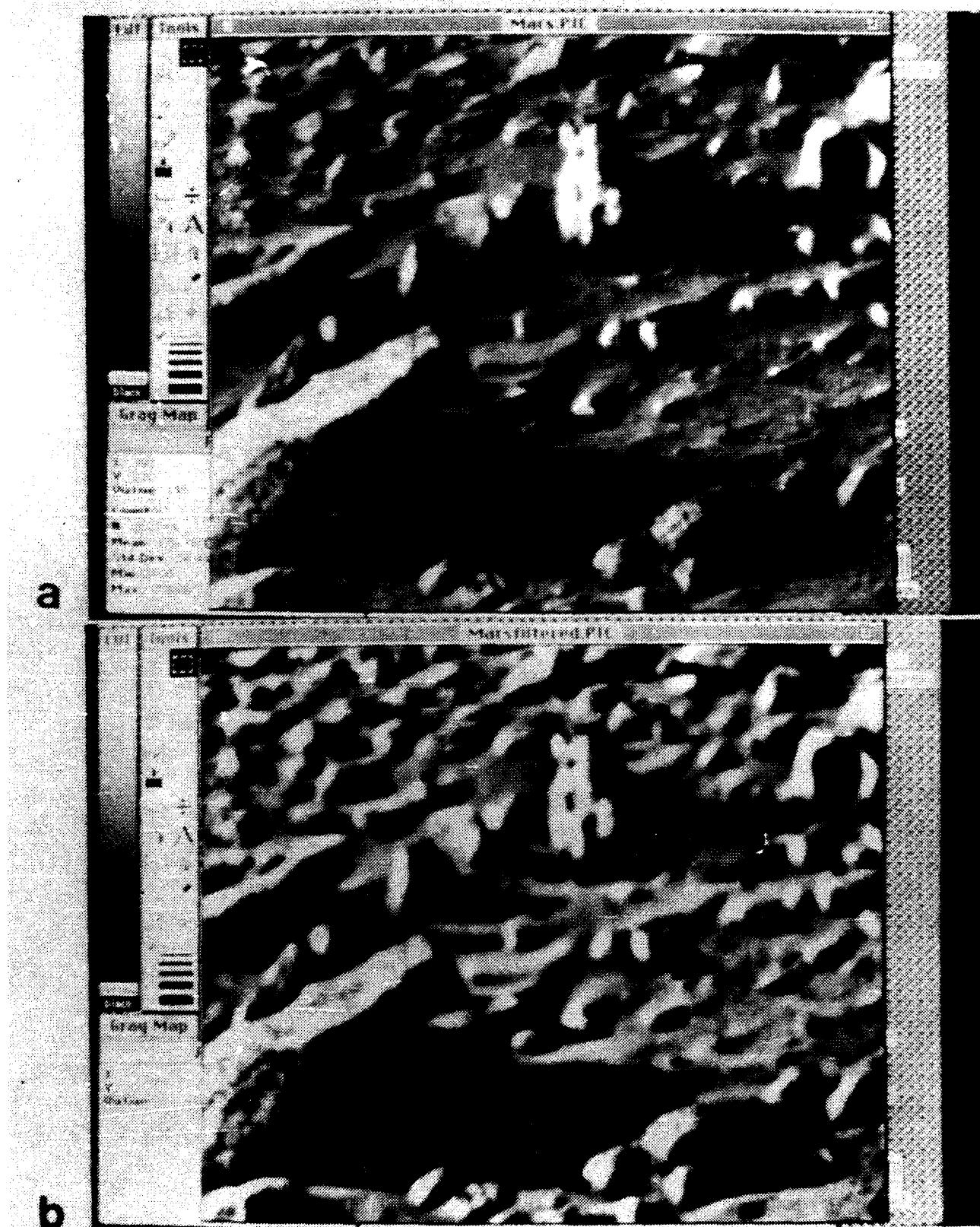
these image enhancement filters do indeed improve object segmentation. This enhancement filtering provides additional object boundaries, defined by shadows, that can be used to match the same object in subsequent scenes, in addition to sharpening up existing object boundaries, providing additional objects in both the foreground and background of the object map. Notice that the small light rock in the bottom right side of the object map in Figs. 3d and 6b is only seen using high contrast boundaries, when the scene is filtered using the image enhancement filter. This small rock is segmented most accurately in the object map, when the parameters are optimized, e.g. the filter size = 7x1, the size threshold = 4 pixels (each segment is at least 4 pixels long), and the amplitude of the sinewave has a peak of at least 5 levels of gray when constructing the horizontal segment map, and 3 levels of gray when constructing the vertical segment map.

The current CVS enables one to test the usefulness of different global and local algorithms on improving object segmentation using modular, efficient software, and by visual inspection of the resultant segment, object, and depth maps in multiple windows, data on the object parameters being printed in a separate window. Once an object has been segmented in the scene, this object can easily be extracted and stored in a separate image file to be used for stimuli in psychophysical and physiological pattern discrimination experiments. The object can be created and its stimulus parameters varied efficiently and systematically using X11-R5 Xwindows primitives, linearizing gamma correction software for increasing the brightness of objects, and optimized C code. Tools to: 1) display the pixel location and grayscale value of each pixel, and 2) zoom or magnify the image centered at any point in the scene, can be used to verify the boundaries and gray level of objects contained in the object map.

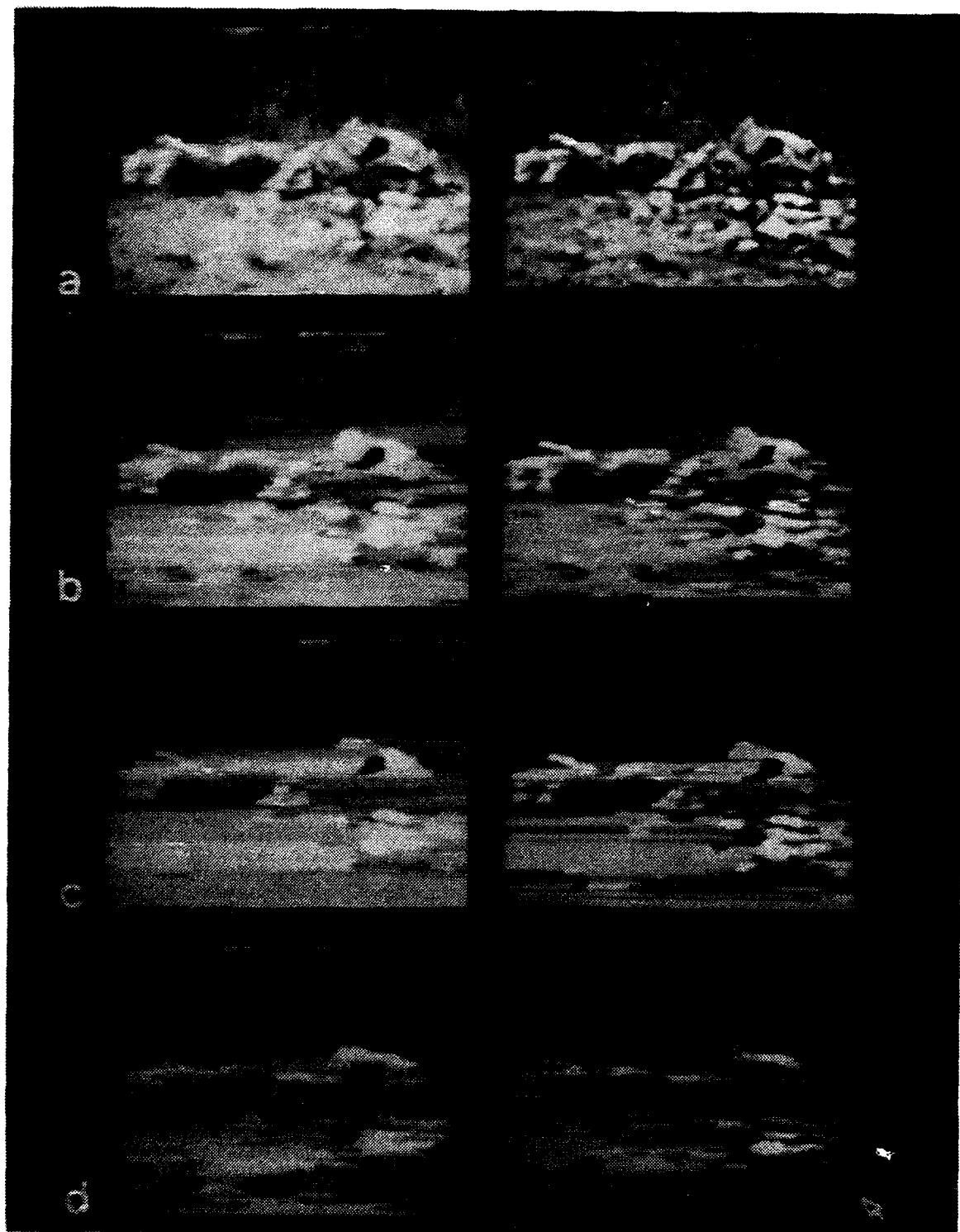
## 6. ROBUSTNESS OF CVS

The success of this method is demonstrated by the speed and robustness of the results when the input consists of natural outdoor scenes -- where the effects of terrain, shadows, scene illumination, reference landmarks, perspective, and scene complexity can be systematically explored. The CVS is only a valuable tool if the algorithms to construct objects can easily be changed to test out: 1) different cooperative analyses to combine line segments when constructing each object, 2) different connection strengths between different object attributes, providing different competitive analyses, and 3) different global and local filtering algorithms to improve object segmentation. Cooperative and competitive analyses are easily changed by modifying algorithms in the segment and object map modules. In so doing, improvements to both the segment and object maps from its original implementation,<sup>33,34</sup> Fig. 1, to its current state, Fig. 3, are easily seen. The type of image enhancement filtering seen on the right side of Figs. 3 and 6, and on the bottom of Fig. 5, show how global scene attributes to reveal objects hidden under shadows also improve object segmentation.

This CVS also uses several ways of improving the robustness of object segmentation using local object attributes. The gain of the boundary detectors can be changed by varying additive constants, such as: 1) the amplitude of the sine detector, 2) the texture threshold, or 3) the size threshold, or by varying 4) the size of the sine and cosine filters. If the filtering parameters are not optimized, as illustrated in Fig. 6, when 1) the sine amplitude for constructing horizontal and vertical segments is doubled, as in 6c, the amplitude of the sinewave has a peak of at least 10 levels of gray when constructing the horizontal segment map, and 6 levels of gray when constructing the vertical segment map, or 2) too large a kernel is used for the size of the sine filters, as in Fig. 6d, where the kernel is a 13 tap filter instead of a 7 tap filter, then the object map construction is noticeably degraded. Although fewer objects are detected when the sine amplitude threshold is doubled, when objects are detected their object boundaries are much more accurate, than when the filter size is doubled. This type of manipulation shows that a multiresolution analysis can be provided by changing an additive constant, instead of changing the size of the filters which significantly slows down the filtering operation by a factor corresponding to the increase in filter size.



**Figure 5.** (a) Video Image of Surface of Mars (Taken From Viking Mars Lander) and (b) Filtered Version of this Scene Using Image Enhancement Filters.



**Figure 6.** (a) Unenhanced (on left side) and Enhanced (on right side) views of same natural scene. (b) Object Maps for best boundary detection filtering parameters, i.e. Filter Size = 7 pixels long, Sine Amplitude = 5 gray levels for horizontal maps, and 3 gray levels for vertical maps, (c) Object Maps for best boundary detection filtering parameters **except** Sine Amplitude = 10 gray levels for horizontal maps, and 6 gray levels for vertical maps, and (d) Object Maps for best boundary detection filtering parameters **except** Filter Size = 13 pixels long.

The filtered images are also noticeably degraded when the minimum size for each segment is reduced from 4 pixels to 1 pixel. Objects in the foreground of the object map are smaller than their actual size and have more jagged edges. In natural scenes where objects have irregularly shaped boundaries, there is a need for extended filters, ones that are more global than pixel-based filters. The need for extended filters is consistent with pattern recognition not being analyzed until visual information reaches the cortex, where the concentric center-surround receptive fields of the retina are expanded into elongated filters in the striate cortex, combining the outputs from at least 48 retinal ganglion cells.<sup>54</sup> These results show the importance of correctly defining the filtering characteristics of the front end of the CVS. Otherwise, a robust object map is not possible, and the depth map cannot be constructed for small or distant objects.

## **7. CONCLUSIONS**

A CVS has been developed that segments objects in natural scenes using algorithms and filtering elements similar to those used by people. Both global scene attributes used to reveal objects masked by shadows, and local object attributes defined by the boundary of contrast differences between an object and its background are used to improve object segmentation. Horizontal and vertical paired odd and even symmetric filters are used to construct a viewer-centered object-based map of the scene that takes into account multiple object attributes. The robustness of object segmentation will be improved by developing better competitive and cooperative algorithms, using data from psychophysical studies of object discrimination in natural scenes to reveal more about the characteristics of cortical feedback. Once object segmentation is much more robust by integrating a dynamic adaptive, multiattribute analysis into one object map, then the next step is to use motion parallax and the redundant object attributes found across subsequent views of the same scene to construct the depth map and remove the effects of occluding objects, extracting an object-centered map from a sequence of viewer-centered object maps.

The CVS I've developed can be used to: 1) generate stimuli for psychophysical and physiological experiments investigating object discrimination in natural scenes, 2) increase the understanding of the algorithms used by normal observers and those with cognitive deficits for discriminating objects in natural scenes, a task requiring multiple cortical areas, 3) reconstruct 3-D object maps of unfamiliar natural scenes, and 4) develop robust and efficient automated pattern recognition systems to aid navigation of both partially sighted human observers and robotic vehicles. To investigate whether cognitive deficits result from integrative or interruptive processing, psychophysical research is needed to investigate both forwards and backwards masking<sup>6</sup> of objects embedded in natural scenes. The psychophysical data will provide normative data for cognitive tasks requiring multiple cortical areas, enabling early detection of cognitive disorders, like dyslexia, that affects at least one in 5 people, and perhaps schizophrenia,<sup>17</sup> as well. By varying the object parameters, such as nonoriented colored objects compared to oriented luminance boundaries used for structure from motion, and instructing the observer to attend to these different object attributes, then the relative contribution of dorsal and ventral pathways for constructing the object and depth maps can be examined. The algorithms derived from this study will provide the basis for developing efficient and robust algorithms for a real-time CVS capable of object discrimination, 3-D perception, and navigation through unknown natural terrain.

## **8. ACKNOWLEDGEMENTS**

This research was supported by Nano Tech Services. I am grateful to Brian Fox for completing the initial implementation in Xwindows and C code at the Biology Division, Caltech, Pasadena, CA in 1990.

## **9. REFERENCES**

1. Alexander, G.E. and Crutcher, M.D. (1990) "Functional architecture of basal ganglia circuits: neural substrates of parallel processing", *Trends in NeuroSciences*, **13**, 266-271.
2. Baizer, J.S., Ungerleider, L.G., and Desimone, R. (1991) "Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques", *J. Neurosci.*, **11**, 168-190.
3. Ballard, D.H. and Brown, C.M. (1982) *Computer Vision*, Prentice-Hall, Englewood Cliffs, N.J.
4. Blakemore, C. and Campbell, F.W. (1969) "On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images", *J Physiol.*, **203**, 237-260.
5. Braddick, O., Campbell, F.W. and Atkinson, J. (1978) "Channels in vision: Basic aspects", in *Handbook of Sensory Physiology VIII: Perception*, eds. R. Held, H. Leibowitz, and H.L. Teuber, Berlin: Springer Verlag.
6. Breitmeyer, B.G. (1984) *Visual Masking: An Integrative Approach*, N.Y. Oxford University Press.
7. Buelthoff, H.H. and Mallot, H.A. (1987) "Interaction of different modules in depth perception", *Proceedings of the First International Conference on Computer Vision, IEEE*, 1987.
8. Castleman, K.R. (1979) *Digital Image Processing*, Prentice-Hall, Englewood Cliffs, N.J.
9. Cavanagh, P. (1987) "Reconstructing the third dimension: Interactions between color, texture, motion, binocular disparity, and shape", *Computer Vision Graphics Image Processing*, **37**, 171-195.
10. Chang, J.J. and Julesz, B. (1984) "Cooperative phenomena in apparent movement perception of random-dot cinematograms", *Vision Research*, **24**, 1781-1788.
11. Daugman, J.G. (1985) "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters", *J. Opt. Soc. Am. A*, **2**, 1160-1169.
12. De Valois, R., Albrecht, D.G. and Thorell, LS. (1982) "Spatial frequency selectivity of cells in macaque visual cortex", *Vision Research*, **22**, 545-559.
13. DeYoe, E.A. and Van Essen, D.C. (1988) "Concurrent processing streams in monkey visual cortex", *TINS*, **11**, 219-226.
14. Felleman, D.J. and Van Essen, DC. (1991) "Distributed hierarchical procesing in the primate cerebral cortex", *Cerebral Cortex*, **1**, 1-47.
15. Gabor, D. (1946) "Theory of communication", *J. IEEE (London)* **93**, 429-457.
16. Graybill, A.M. (1990) "Neurotransmitters and neuromodulators in the basal ganglia", *Trends in Neurosciences (TINS)*, **13**, 244-254.
17. Green, M, and Walker, E. (1984) "Symptom correlates of vulnerability to backward masking in schizophrenia", *Am. J. Psychiatry*, **143**, 181-186.
18. Gross, C.G., Rocha-Miranda, C.E. and Bender, D.B. (1972) "Visual properties of neurons in the inferotemporal cortex of the macaque", *J Neurophysiol.*, **35**, 96-111.
19. Grossberg, S. and Mingolla, E. (1985b) "Neural Dynamics of Perceptual grouping: Texture, boundaries, and emergent segmentation", *Perception & Psychophysics*, **38**, 141-171.
20. Helmholtz, H. von (1925) *Helmholtz's Treatise on Physiological Optics*, ed. J.P.C. Southall, Dover Pub, N.Y.
21. Horn, B.K.P. (1986) *Robot Vision*, MIT Press, McGraw-Hill Book Co., Cambridge, Ma..
22. Hubel, D.H. and Livingstone, M.S. (1990) "Color and contrast sensitivity in the lateral geniculate body and primary visual cortex of the macaque monkey", *J. Neurosci.* **10**, 2223-2237.
23. Hubel, D.H. and Wiesel, T.N. (1968) "Receptive fields and functional architecture of monkey striate cortex", *J. Physiol.*, **195**, 215-243.

24. Julesz, B. (1971) *Foundations of Cyclopean Perception*, University of Chicago Press, Chicago.

25. Koffka, K. (1935) *Principles of Gestalt Psychology*, Harcourt, Brace and World, NY.

26. Kulikowski J.J., Marcelja S., Bishop P.O. (1982) "Theory of spatial position and spatial frequency relations in the receptive fields of simple cells in the visual cortex", *Biol. Cybern.* **43**, 187-198.

27. Lawton, T.B. (1984) "The effect of phase structures on spatial phase discrimination", *Vision Research*, **24**, 137-148.

28. Lawton, T.B. (1985) "The spatial frequency spectrum of a pattern changes the visibility of spatial phase differences", *J. Optical Society of Am. A*, **2**, 1140-1152.

29. Lawton, T.B. (1988) "Improved word recognition for observers with age-related maculopathies using compensation filters," *Clinical Vision Sciences* , **3**, 125-135.

30. Lawton, T.B. (1988) "When perceived velocity is independent of contrast", being reviewed by *Visual Neuroscience*; research originally presented at the 11th European Conf. on Visual Perception, Bristol, England.

31. Lawton, T.B. (1989) "Improved reading performance using individualized compensation filters for observers with losses in central vision", *Ophthalmology*, **96**, 115-126.

32. Lawton T.B. (1989) "Outputs of paired Gabor filters summed across the background frame of reference predict the direction of movement" *IEEE Transactions in Biomedical Engineering* (Feature Issue on Neural Systems and Neural Engineering) **36**, 130-139.

33. Lawton, T.B. (1990) "Neural algorithms for automated pattern recognition in natural scenes", Invited Paper in *Proceedings from IEEE Transactions on Parallel Processing* , in Fullerton, CA, April 4, **2**, 4-10, where chaired session on Neural Network Applications.

34. Lawton, T.B. (1990) "Determining Sense of Motion in Robotic Vision", New Technology Report No. 17552, *NASA Tech Briefs*, **14:4** , 84-85.

35. Lawton, T.B. (1992) "Image Enhancement Filters Significantly Improve Reading Performance" *Ophthal. and Physiol. Optics*, **12**, 193-200.

36. Longuet-Higgins, H.C. and Prazdny, K. (1980) "The interpretation of a moving retinal image", *Proc. R. Soc. Lond. B.* **208**, 385-397.

37. Marcelja, S. (1980) "Mathematical description of the response of simple cortical cells" *J. Optical Soc. Am.*, **70**, 1297-1300.

38. Marr, D. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, San Francisco: Freeman (1982).

39. Marshall, J. (1990) "Self-organizing neural network for the perception of visual motion", *Neural Networks*, **3**, 45-74.

40. Maunsell, J.H.R. and Van Essen, D.C. (1983) "The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey", *J. Neurosci.*, **3**, 2563-2586.

41. Maunsell, J.H.R. and Newsome, W.T. (1987) "Visual processing in monkey extrastriate cortex", *Ann. Rev. Neuroscience* **10**, 363-401.

42. McKee, S.P., Silverman, G.H., Nakayama, K. (1986) "Precise velocity discrimination despite random variations in temporal frequency and contrast", *Vision Res.*, **26**, 609-620.

43. Merigan, W.H. and Maunsell J.H.R. (1993) "How parallel are the primate visual pathways?", *Annu. Rev. Neurosci.* **16**, 369-402.

44. Morel, A. and Bullier, J. (1990) "Anatomical segregation of 2 cortical visual pathways in the macaque monkey", *Vis. Neurosci.*, **4**, 555-578.

45. Movshon, J.A. , Thompson, I.D. and Tolhurst, D.J. (1978) "Spatial summation in the receptive fields of simple cells in the cat's striate cortex", *J. Physiol.* **283**, 53-77.

46. Nelson, J.I., (1985) "The cellular basis of perception", in *Models of the Visual Cortex*, eds. D. Rose and V.G. Dobson, John Wiley and Sons, New York, 108-122.

47. Newsome, W.T., Wurtz, R.H., and Komatsu, H. (1988) "Relation of cortical areas MT and MST to pursuit eye movements II. Differentiation of retinal from extraretinal inputs", *J Neurophysiol.* **60**, 604-620.

48. Ono, H., Roger, B.J., Ohmi, M. and Ono, M.E. (1987) "Dynamic occlusion and motion parallax in depth perception", *Perception*, **17**, 255-266.

49. Perrett, D.I., Oram, M.W., Harries, M.H., Bevan, R., Hietanen, J.K., Benson, P.J. and Thomas, S. (1991) "Viewer-centered and object-centered coding of heads in the macaque temporal cortex", *Exp. Brain Res.* **86**, 159-173.

50. Pollen, D.A., Ronner, S.F. (1981) "Phase relationships between adjacent simple and complex cells in the visual cortex of the cat", *Science*, **212**, 1409-1411.

51. Posner, M.I. and Peterson, S.E. (1990) "The attention system of the human brain", *Annual Review of Neuroscience*, **13**, 25-42.

52. Robbins, T.W. and Everitt, B.J. (1988) "Psychopharmacological studies of arousal and attention", in *Cognitive Neurochemistry*, eds. Stahl, S.M., Iverson, S.D. and Goodman, E.C., Oxford University Press, 135-170.

53. Robson, J.G. (1975) "Receptive fields: Neural representation of the spatial and intensive attributes of the visual image" in *Handbook of Perception*, eds. E.C. Carterette and M.P. Friedman, **5**, 81-112.

54. Robson, J.G. (1993) "Low level representations in the cortex", *Proceedings of Computational Vision Based on Neurobiology*, sponsored by SPIE, Asilomar, CA July 6-9.

55. Rockland, K.S. and Pandya, D.N. (1979) "Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey", *Brain Res.* **179**, 3-20.

56. Sakitt, B. and Barlow, H.B. (1982) "A model for the economical encoding of the visual image in cerebral cortex", *Biol. Cybernet.*, **43**, 97-108.

57. Tanaka, K., Saito, H., Fukada, Y., and Moriya, M. (1991) "Coding visual images of objects in the inferotemporal cortex of the macaque monkey", *J. Neurophysiol.* **66**, 170-189.

58. Treisman, A. (1988) "Features and objects: The fourteenth Bartlett Memorial lecture", *Quart. J. Exp. Psych.*, **40A**, 201-237.

59. Treisman, A. and Gelade, G.A. (1980) "A feature-integration theory of attention", *Cognitive Psychology*, **12**, 97-136.

60. Ungerleider, L.G. and Mishkin, M. (1982) "Two cortical visual systems", in: *Analysis of Visual Behavior*, eds. D.J. Ingle, M.A. Goodale and R.J.W. Mansfield, MIT Press, Cambridge, MA, 549-586.

61. Van Essen, D.C. and Maunsell, J.H.R. (1983) "Hierarchical organization and functional streams in the visual cortex", *Trends in Neurosciences*, **6**, 370-375.

62. Wurtz, R.H., Yamasaki, D.S., Duffy, C.J. and Roy, J.P. (1990) "Functional specialization for visual motion processing in primate cerebral cortex", *Cold Spring Harbor Symposia Quantitative Biology*, **105**, 717-727.

63. Young, M.P. (1993) "The connectional organization of neural systems in the primate cerebral cortex", *Proc. Royal Society: Biol. Studies*, **252**, 13-18.

64. Young, M.P. and Yamane, S. (1992) "Sparse population coding of faces in the inferotemporal cortex", *Science*, **256**, 1327-1331.

65. Zeki, S. and Shipp, S. (1988) "The functional logic of cortical connections", *Nature*, **335**, 311-317.

## **The influence of figural interpretation on the selective integration of visual motion signals**

Thomas D. Albright and Gene R. Stoner

The Salk Institute for Biological Studies  
La Jolla, CA 92037

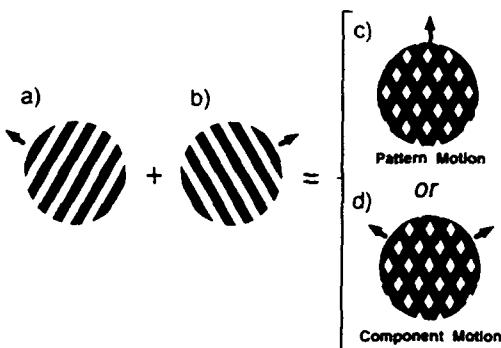
### **ABSTRACT**

The solution to the computational problem of reconstructing object motion from retinal image motion is underconstrained. In an effort to converge on a solution to this problem, the primate visual system appears to rely upon image cues that lead to an interpretation of the spatial relationships between objects in a visual scene. Psychophysical experiments illustrate this phenomenon through the apparent dependence of motion signal integration on luminance-based cues for occlusion and perceptual transparency. Neurophysiological studies of the cell populations thought to underlie motion signal integration reveal a change in directional selectivity that precisely parallels the perceptual phenomenon. Among obstacles faced in attempts to understand the neural bases of primate vision, the integration of motion signals holds a unique position: The computational problem is well-defined, a specific neural substrate has been identified, and the solution to the integration problem is absolutely critical for visually-guided behavior. As such, it stands as a model system for exploring the relationships between neuronal phenomena, perception, and behavior.

### **1. MOTION SIGNAL INTEGRATION**

The motions of objects in the world often give rise to a complex pattern of moving and overlapping features in the retinal image. From such intangibles it is clearly possible for the primate visual system to construct a veridical representation of moving objects. Because the solution is otherwise grossly underconstrained, we have proposed that this process relies upon tacit knowledge of the "rules" by which two-dimensional (2D) retinal image features are formed from their real-world 3D counterparts<sup>1,2</sup>. Such information is essential for perceptual interpretation of the spatial relationships between moving image features, which in turn allows moving features to be integrated according to object of origin.

This hypothesis regarding the integration of visual motion signals can be readily tested in psychophysical and neurophysiological experiments using stimuli that have been termed "moving plaid patterns"<sup>3,4</sup>. These 2D patterns are formed, as illustrated in Figure 1, by superimposition of two overlapping and drifting 1D gratings. Plaids provide a simple laboratory counterpart to real-world situations that give rise to overlapping contours in the retinal image. Their value in this context comes from the fact that under some conditions the two grating components are seen to move independently or "non-coherently", while under other conditions the two components are seen to form part of a single 2D pattern that moves "coherently". By manipulating various image parameters it becomes possible to identify the conditions that lead to these two different percepts. Our hypothesis predicts that these conditions will correspond to those that influence perceptual parsing of the image into two surfaces vs. one.



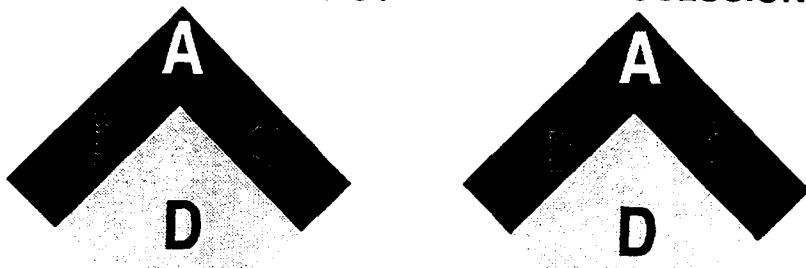
**Figure 1:** Moving plaid patterns are produced by superimposition of two drifting periodic grating. The resultant percept is either that of a coherently moving two-dimensional plaid pattern or two one-dimensional gratings sliding past one another, depending on a variety of stimulus parameters.

## 2. PSYCHOPHYSICAL STUDIES OF MOTION SIGNAL INTEGRATION

Using drifting plaid patterns as stimuli, it was originally shown that the likelihood of perceptual motion coherence decreases when the component gratings differ significantly along the dimensions of spatial frequency or luminance contrast<sup>3</sup>. Subsequent psychophysical experiments demonstrated that components having different binocular disparities (thus appearing to lie in different depth planes) are also less likely to cohere<sup>5</sup>, as are those created by modulation along different color-opponent axes<sup>6</sup>. These observations have typically been explained by invoking relatively simple channel-based mechanisms for selective integration<sup>7</sup>. The observations are nonetheless consistent with our functional proposal, whereby motion signal integration hinges upon determination of the figural origins of moving image features.

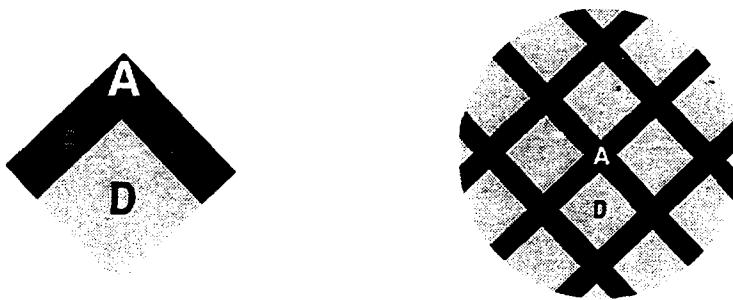
To explore this possibility more extensively, we chose to manipulate luminance cues that directly influence perceived depth ordering of surfaces<sup>1</sup>. The appropriateness of these cues (and their ubiquity in natural images) can be evaluated by considering how retinal images are formed. It is often the case that when one moving object passes in front of another, the nearer object occludes the distant object. At the point of overlap the luminance may be exclusively that of the nearer surface. In other instances, a *transparent* foreground object may attenuate, but not occlude, light from the distant surface. A special case of such attenuation is that characteristic of shadows. Bearing in mind these properties of image formation, there are some elementary luminance relationships that dictate whether simple patterns, such as those shown in Figure 2, are physically consistent with two overlapping surfaces or four distinct surfaces. These luminance relationships are captured by the "rules" of perceptual transparency<sup>8,9</sup>. Accordingly, luminance relationships falling between the extremes of occlusion and shadow-like transparency can arise from independent but spatially overlapping surfaces. Luminance variations occurring at the region where two independent surfaces intersect (region A in Figure 2) are considered "extrinsic"<sup>10</sup>, i.e., a consequence of the figural relationships between surfaces rather than a property of the surfaces themselves. Luminance variations that lie outside of this occlusion-transparency range cannot be attributed to object interrelationships and must, therefore, result from "intrinsic" surface properties, such as differences in surface reflectance. In accordance with our hypothesis, we predicted that the figural relationships implied by these luminance variations would gate the selective integration of motion signals.

## "PURE" TRANSPARENCY      "PURE" OCCLUSION

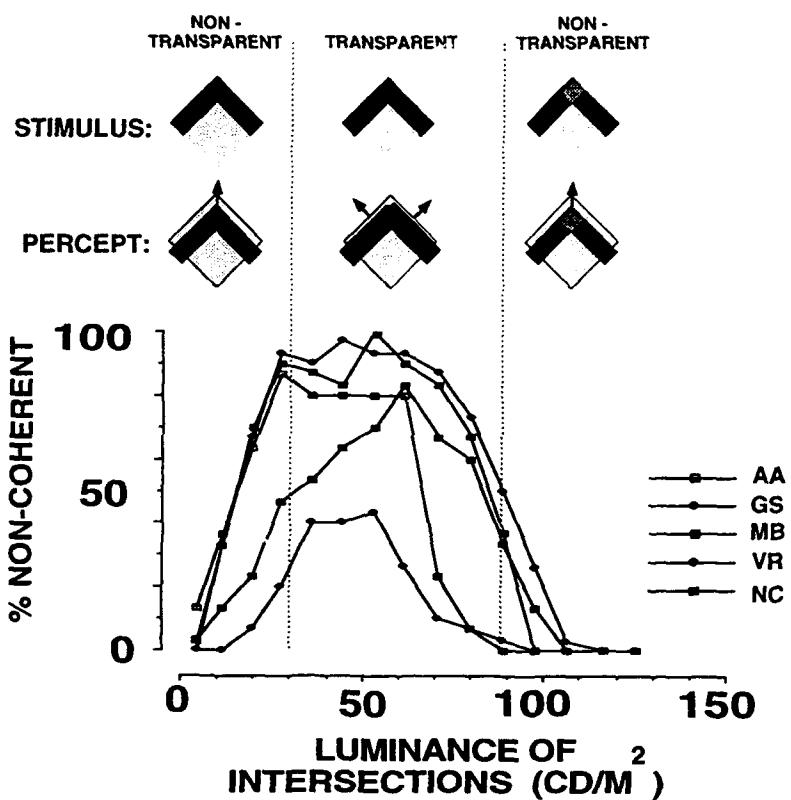


**Figure 2:** The procedures for creating perceptually transparent plaid patterns are derived from the physics of transparency<sup>19</sup>. Simply put, luminance ratios within the pattern must be physically consistent with the transmittance of light from a far surface *through* a near surface. Appropriate luminance ratios convey a sense of depth ordering (and hence image segmentation) in a pattern devoid of other depth cues. The zone of perceptual transparency is bounded by two extremes: "pure" transparency and "pure" occlusion. Pure transparency (left) occurs when the near transparent surface AC reflects no light but transmits light from the surface behind it. Pure occlusion (right) occurs when the near surface AC reflects light but transmits no light from the surface behind it. From Stoner and Albright<sup>2</sup>.

This prediction was tested using plaid patterns constructed with reference to the laws of perceptual transparency (Figure 3). As predicted, when the luminance relationships were configured such that the component gratings appeared occlusive or transparent, human subjects generally reported a percept of non-coherent motion (i.e., the two gratings appeared to slide across one another) (Figure 4). Alternatively, when the luminance configuration was incompatible with transparency or occlusion, subjects generally reported a percept of coherent motion<sup>1</sup>.



**Figure 3:** Each plaid can be viewed as a tessellated image composed of four distinct repeating subregions, identified as A, B, C, and D. Region D is normally seen as background. Regions B and C are seen as narrow overlapping surfaces, and the remaining region A is seen as their intersection. Perceptual transparency was manipulated in both psychophysical<sup>1</sup> and neurophysiological<sup>17</sup> experiments by adjusting the luminance of region A, while the luminances of regions B, C, and D were held constant.



**Figure 4:** Results from psychophysical experiments examining the effects of perceptual transparency on motion coherency. Probability of the component motion percept is plotted as a function of the intersection luminance for appropriately configured plaid patterns (see Figures 2 and 3). Both gratings were of the same spatial frequency (1.75 cyc/°). On each trial the individual gratings were moved at an angle of 135° relative to one another at a speed of 3°/s, resulting in a pattern speed of 8°/s. Pattern direction was either up or down, and varied on a random schedule. Intersection luminance was varied in equal steps, such that it was either compatible or incompatible with transparency. The "transparency zone" extends from pure (multiplicative) transparency (35 cd/m<sup>2</sup>) up to the point of occlusion (90 cd/m<sup>2</sup>). A percept of non-coherent component motion is most likely within a region roughly centered on the transparency zone. Each data point represents the mean of 30 trials for each intersection luminance value. Data are shown for five subjects. Adapted from Stoner et al.<sup>1</sup>.

These results support our general hypothesis regarding the contribution of image segmentation cues to motion signal integration. They tell us little about the mechanism involved, however. One common proposal<sup>2,11,12</sup> is based upon the fact that the luminance manipulations involved in simulating transparency and occlusion also vary the strength of fourier components that move in the coherent direction. Indeed, if one allows for an early logarithmic signal compression<sup>13</sup>, the resultant strength of such phantom fourier components roughly accounts for the results of Stoner et al<sup>1</sup>. This low-level explanation is called into question, however, by the results of other recent experiments, which show that the perception of transparency -- and, in turn, motion coherence -- is also dependent upon image cues that do not affect the strength of phantom fourier components. For example, perceptual transparency is inherently dependent upon figural cues that influence the visual system's ability to interpret the relationship between foreground and background in a visual scene. The percept of transparency in the left panel of Figure 2 relies upon the fact that the observer interprets surface *AC* as foreground and surface *BD* as the unattenuated background. Clearly, the converse figural interpretation (which can be willed with some effort) does not lead to a percept of transparency. The explanation for this phenomenon is rooted in the fact that transparent surfaces typically do not

enhance the contrast of surfaces seen through them. Hence, the luminance variations in surface *BD* cannot arise solely by virtue of it being transparent and overlying surface *AC*.

To further explore this phenomenon and examine its influence over motion signal integration, Stoner and Albright<sup>14</sup> used pictorial cues to bias foreground interpretation. Two methods were used (Figure 5). Human subjects viewed plaid patterns for which foreground/background interpretation was manipulated and, as in the original transparency experiments, they reported whether they perceived coherent or non-coherent motion. Only the luminance of one region of the pattern was varied (region *A* in Figure 5). Using the means indicated, foreground/background interpretation was manipulated such that region *A* was likely to be perceived as either (1) the intersection of the two component gratings, or (2) the unobstructed background. As expected, both the percept of transparency and motion coherence were heavily dependent upon foreground/background interpretation, as manipulated by either technique. This result is not compatible with explanations based upon simple detection of the motions of phantom fourier components. Although the details of an alternative mechanism have yet to be worked out, the result adds further weight to our claim that motion integration has access to image segmentation processes built upon the rules governing retinal image formation from natural scenes.

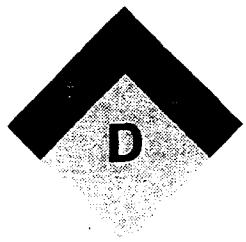
### 3. NEUROPHYSIOLOGICAL STUDIES OF MOTION SIGNAL INTEGRATION

The conceptual framework described above implies the existence of at least two stages of motion processing in the primate brain, which provide the roles of motion detection and integration, respectively. Neurophysiological studies employing plaid patterns as visual stimuli have allowed a tentative identification of the neuronal populations corresponding to these two stages. Movshon and colleagues<sup>7</sup> examined the directional selectivity of V1 neurons to perceptually coherent plaid patterns. Consistent with their orientation tuning, V1 neurons were found to signal only the motion of the 1D components. Such neurons have been referred to as "component type" and are believed to represent the first motion processing stage. The integration process appears to take place in the middle temporal visual area (area MT), an area that receives direct input from V1 and is thought to play a crucial role in motion processing<sup>15</sup>. While many MT neurons (40%) also appear to be component type when tested under these conditions, a small population (25%) respond in a way that reflects sensitivity to pattern motion<sup>7,16</sup>. Neurons of this latter type have been called "pattern type" and are presumed to constitute the second stage of motion processing, at which motion signal integration takes place (Figure 6).

Although the factors affecting motion signal integration have been studied in some detail psychophysically, until recently virtually nothing was known of the neural interactions underlying these effects. Previous studies that attempted to classify directionally selective neurons on the basis of responses to component or pattern motion used plaid patterns that were *always perceptually coherent*. Since these neurons are believed to play some significant role in the integration process, we hypothesized that their behavior would be altered by stimulus attributes known to influence perceptual integration of motion signals. As in our earlier psychophysical experiments described above<sup>1</sup>, perceptual motion coherence was manipulated by altering the stimulus conditions such that plaid patterns were either consistent or inconsistent with transparency. Directional selectivity of single MT neurons was assessed using each of three different plaid configurations<sup>17</sup>. Two of these stimuli elicited a percept of coherent pattern motion; the third elicited a percept of independently moving components. Data obtained from a typical neuron are illustrated in Figure 7. When stimulated using either of the perceptually coherent plaids, this cell responded more strongly when the 2D pattern moved in the neurons's preferred direction than when either of the 1D components moved in that same direction. As can be seen in Figure 6, this type of tuning is characteristic of pattern type neurons<sup>7,16</sup>. When stimulated using the transparent and perceptually non-coherent plaid, however, this cell's behavior underwent a marked transformation: the pattern response dropped while component responses became elevated. The resultant bilobed directional tuning curve is characteristic of component type neurons (Figure 6). As was the case for the majority of neurons in our sample, this cell's sensitivity to component motion increased (and sensitivity to pattern motion decreased) when the visual stimulus was configured such that a percept of component motion became more likely.

LUMINANCE CONTRAST VIEWED THROUGH  
TRANSPARENT FOREGROUND SURFACE [AC] SHOULD BE  
LESS THAN THAT OF BACKGROUND [BD]

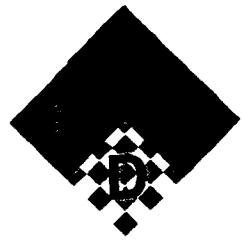
PERCEPTUALLY  
TRANSPARENT



PERCEPTUALLY  
NON-TRANSPARENT

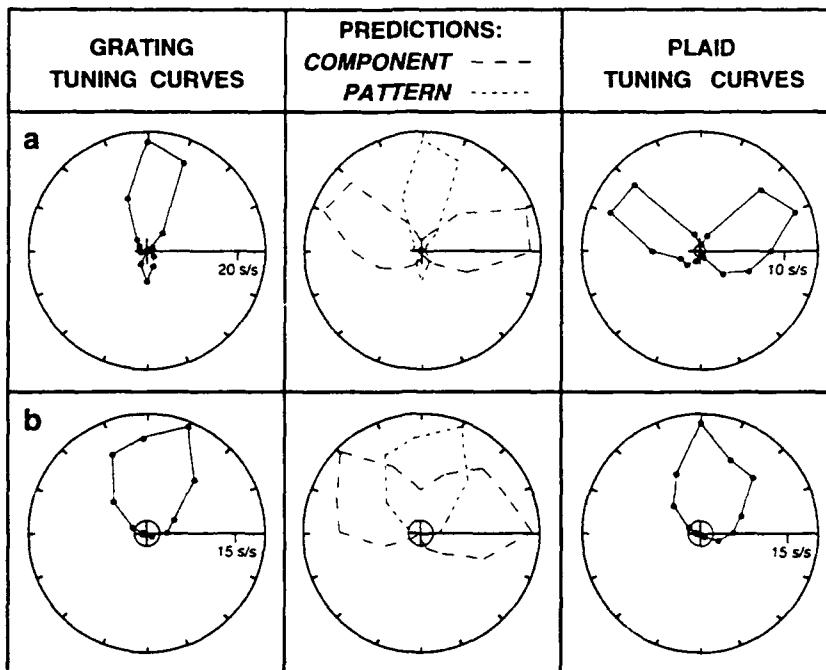


FOREGROUND CONTRAST (A/C)  
IS LESS THAN  
BACKGROUND CONTRAST (B/D)



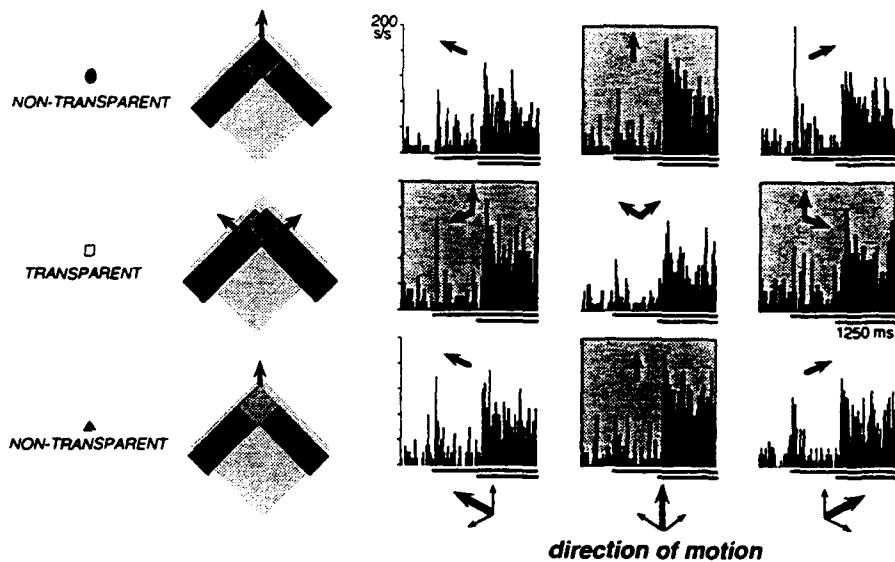
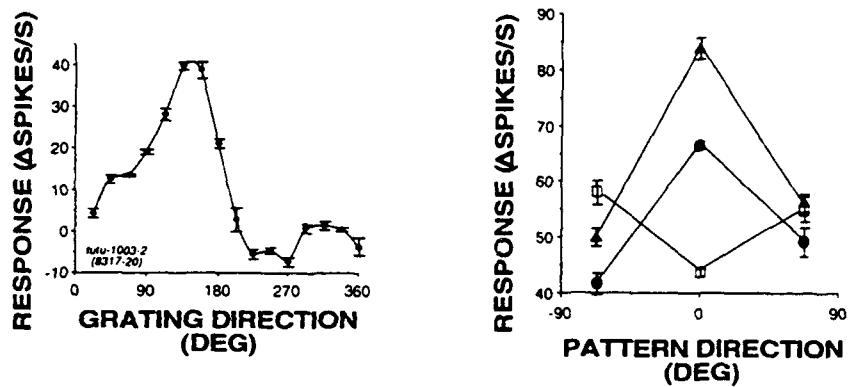
FOREGROUND CONTRAST (B/D)  
IS GREATER THAN  
BACKGROUND CONTRAST (A/C)

**Figure 5:** Schematic illustration of two methods used by Stoner and Albright<sup>14</sup> to manipulate the perception of foreground and background in plaid patterns. The plaids were tessellated versions of these basic patterns (see Figure 3). One method, depicted in the top row, involved manipulating the relative sizes of the plaid sub-regions. The larger regions (region D on the left and region A on the right) are usually seen as background<sup>18,19</sup>. A second method, shown in the bottom row, was to place a static checkerboard in the putative background region. The plaid motion progressively occluded/disoccluded this pattern, causing the textured region (region D on the left and region A on the right) to be seen as background (note that the "occlusion" of this checkerboard by the putatively transparent gratings is physically consistent with the contrast reduction associated with transparent surfaces or the blurring common with translucent surfaces such as smoked glass). Both methods reliably influence perceptual assignment of foreground/background, while leaving the space-averaged luminance of the four regions constant. The reversal of foreground/background assignment, in turn, had a profound effect on both perceptual transparency and motion coherency judgements by human observers. From Stoner and Albright<sup>14</sup>.



**Figure 6:** Data from two MT neurons representing "component" (**a** row) and "pattern" (**b** row) stages of motion processing in cortical visual area MT or the rhesus monkey. Direction tuning curves were acquired using a drifting sine-wave grating (first column) or a perceptually coherent plaid pattern (third column). Responses elicited by each stimulus type, moving in each of 16 different directions, are plotted in a polar format. The radial axis represents response amplitude (s/s = mean spike rate during presentation of the stimulus within the receptive field), the polar angle corresponds to the direction of motion, and the small circle in the center of each polar plot represents the level of spontaneous activity. Both cells exhibit a single peak in the grating tuning curve. From these curves, responses to the moving plaid pattern were predicted in accordance with either component or pattern assumptions (second column). The component predictions reflect sensitivity to both oriented components in the plaid pattern. The pattern predictions reflect sensitivity to the composite appearance of the plaid. By definition, the behavior of the component motion neuron conforms to the component prediction while that of the pattern motion neuron conforms to the pattern prediction. Adapted from Rodman and Albright<sup>16</sup>.

These neurophysiological data demonstrate that visual stimulus conditions that influence image segmentation and perceptual motion coherence, also lead to systematic differences in the directional tuning of the neurons that are believed to underlie motion signal integration. Although the neuronal circuits responsible for this perceptually-dependent gating of directional selectivity are entirely unknown, these new results suggest that image segmentation signals might interact with motion signals at the integration stage.



**Figure 7:** Neural correlates of perceptual motion signal integration. **A:** Differential responses of an MT pattern-type neuron to coherent vs. non-coherent plaid patterns. Directional tuning for a single drifting grating is plotted at left. Responses to coherent and non-coherent plaids are plotted at center. When stimulated with coherent plaid patterns, response was maximal when the pattern moved in the neuron's preferred direction ( $0^\circ$  in graph at center; highlighted histograms in top and bottom rows at right). However, when stimulated with non-coherent plaid patterns, responses were maximal when either component moved in the preferred direction ( $\pm 67.5^\circ$  in graph at center; highlighted histograms in center row at right). Adapted from Stoner and Albright<sup>17</sup>.

#### 4. ACKNOWLEDGEMENTS

The research reported herein was supported in part by a grant from the National Institute of Mental Health and a Development Award from the McKnight Endowment fund for Neuroscience. Gene Stoner was partially supported by a Research Fellowship from the McDonnell-Pew Center for Cognitive Neuroscience at San Diego. We thank J. Costanza for superb technical assistance.

## 5. REFERENCES

1. Stoner, G.R., Albright, T. D., and Ramachandran, V. S. Transparency and coherence in human motion perception. *Nature*. **344**: 153-155, 1990.
2. Stoner, G.R. and Albright, T.D. Image segmentation cues in motion processing: Implications for modularity in vision. *J. Cog. Neurosci.*, **5**(2):129-149, 1993.
3. Adelson, E.H. and Movshon, J.A. Phenomenal coherence of moving visual patterns. *Nature*. **300**:523-525, 1982.
4. De Valois, K.K., De Valois, R.L., and Yund, E.W. Responses of striate cortex cells to grating and checkerboard patterns. *J. Physiol. (Lond.)*, **291**:483-505, 1979.
5. Adelson, E.H. Binocular disparity and the computation of two-dimensional motion. *J. Opt. Soc. Am. A*, **1**:1266, 1984.
6. Krauskopf, J. and Farell, B. Influence of colour on the perception of coherent motion. *Nature*. **348**:328-331, 1990.
7. Movshon, J.A., Adelson, E.A., Gizzi, M. and Newsome, W.T. The analysis of moving visual patterns. In (Eds.) C. Chagas, R. Gattass and C. G. Gross. *Study Group on Pattern Recognition Mechanisms*, Pontifical Academia Scientiarum: Vatican City pp. 117-151, 1985.
8. Metelli, F. The perception of transparency. *Scientific American*, **230**: 91-95, 1974.
9. Beck, J., Prazdny, K. and Ivry, R. The perception of transparency with achromatic colors. *Percept. Psychophysics*. **35**:407-422, 1984.
10. Shimojo, S., Silverman, G.H. and Nakayama, K. Occlusion and the solution to the aperture problem for motion. *Vision Res.* **29**:619-626, 1989.
11. Mulligan, J.B. Non-linear combination rules and the perception of visual motion transparency. *Vision Res.* **33**:2021-2030, 1992.
12. Trueswell, J.C., and Hayhoe, M.M. Surface segmentation mechanisms and motion perception. *Vision Res.* **33**:313-328, 1993.
13. MacLeod, D.I.A. Visual sensitivity. *Annual Review of Psychology*. **29**:613-645, 1978.
14. Stoner G.R. and Albright T.D. The influence of foreground/background assignment on transparency and motion coherence in plaid patterns. In preparation, 1994.
15. Albright, T.D. Cortical processing of visual motion. In F.A., Miles and J., Wallman (Eds.), *Visual Motion and its Role in the Stabilization of Gaze*. Amsterdam: Elsevier, pp.177-201, 1993.
16. Rodman, H.R. and Albright, T.D. Single-unit analysis of pattern-motion selective properties in the middle temporal visual area (MT). *Exp. Brain Res.* **75**:53-64, 1989.
17. Stoner, G.R. and Albright, T.D. Neural correlates of perceptual motion coherence. *Nature*. **358**:412-414, 1992.
18. Petter, G. Nuove ricerche sperimentali sulla totalizzazione percettiva. *Rivista di psicologia*, **50**: 213-227, 1956.
19. Vallortigara, G. and Bressan, P. Occlusion and the perception of coherent motion. *Vision Res.* **31**:1967-1978, 1991.

**WORKSHOP**  
**COMPUTATIONAL VISION BASED ON NEUROBIOLOGY**

**Wednesday, July 7, 1993**

**Moderated by Drs. Malcolm Young and Teri Lawton**

**Participants:**

Dr. Malcolm Young, University Laboratory of Physiology, Parks Road, Oxford University, Oxford, England.  
Professor Jennifer Lund, Department of Visual Science, Institute of Ophthalmology, University of London, England.  
Dr. Teri Lawton, Nano Tech Services, 3700 Peninsula Rd., Channel Islands Harbor, CA. and Doheny Eye Institute, 1450 San Pablo St., USC, Los Angeles, CA.  
Professor John Robson, Physiological Laboratory, Downing Street, Cambridge University, Cambridge, England.  
Professor Ralph Siegel, Department of Molecular Biology, Rutgers University, Newark, N.J.  
Professor A.B. Bonds, Department of Electrical Engineering, Vanderbilt University, Nashville, TN.  
Professor Stan Klein, School of Optometry, University of California, Berkeley, CA.  
Dr. Christopher Tyler, Smith Kettlewell Eye Research Institute, 2232 Webster St, San Francisco, CA  
Professor Peter Lennie, Center for Visual Sciences, University of Rochester, New York.  
Professor Lucia Vaina, Department of Biomedical Engineering, Boston University, MA.  
Jeff Teeters, Student of Frank Werblin, University of California, Berkeley, CA.  
Professor Russell De Valois, Psychology Department and Visual Science Group, University of California, Berkeley, CA.  
Dr. Izumi Ohzawa, School of Optometry, University of California, Berkeley, CA.  
Dr. Michael Oram, School of Psychology, University of St. Andrews, St. Andrews, Scotland.  
Professor Anne Treisman, Psychology Department, University of California, Berkeley, CA.  
Dr. Leslie Ungerleider, Laboratory of Neuropsychology, NIMH, NIH, Bethesda, MD.

**Malcolm Young:** What is the importance of using neurobiology for developing robust computational vision models? How many of you are modelers? (A few hands are raised)

**Jenny Lund:** Models help us to do robust biology. Neurobiology needs modelers.

**Malcolm Young:** What about the opposite, can biology help us develop robust computational vision systems?

**Teri Lawton:** Biological systems have evolved to be able to solve pattern recognition in natural scenes. The only hope we have to home in on the answers is to develop robust computational vision systems. We must use what we know about neurobiology to upgrade these models. Once you get to the cortex where there are multiple areas, these different regions communicating using feedforward and feedback connections, then it becomes extremely complicated. The only hope we have to figure out what's really going on is to start putting these models into the computer, and testing out their predictions. The models have become too complex to analyze in our heads. Even if we only implement the limited algorithms that have been revealed by neurobiology in our computational vision systems, as shown by the system I developed, we have a system far more advanced than any other system developed by the experts in computer vision. It is nice when your research can lead to a product that lightens up the task of implementing and testing out different models for pattern discrimination and recognition.

In the models I am developing, we start with what are the algorithms that are used for a particular task. Then we don't need to implement a computational vision system that is as expansive as the traditional neural network models where individual cells, instead of algorithms that represent the function of a group of cells, are used as the basic filtering elements. There is no way that we can duplicate the anatomical complexity of the human visual system using the computer. I think that we can only get a snapshot, an extraction of some subset of the modules in the visual system, those modules used for the task being modeled. We should use computational visual systems to understand how the visual system functions.

**John Robson:** I can't understand what you're saying. Why do you think the visual system is complex? What do you mean by saying it's complex? We don't understand it.

**Teri Lawton:** It has many different visual areas that have many different connections between them.

**John Robson:** We only have 35 or so.

**Ralph Siegel:** If the visual system was simple, then we'd understand it.

**John Robson:** No. I don't think that follows at all. I think that the visual system might be very simple, and we still might not understand it. We just don't know how to look at it or how to talk about it, that's all.

**Ralph Siegel:** Look at all the smart people who have studied vision. Can they all be that stupid?

**John Robson:** Yes. (Floor-Yes. Followed by laughter) Basically, none of us know what we're doing. (More laughter from the Floor. )

**Teri Lawton:** Yes, but that's why we need computers to help us to get these answers.

**A.B. Bonds:** Oh, No. (Laughter from the Floor)

**Teri Lawton:** I think that all of the people here have incredibly good insights into what's going on, and they've been getting some very interesting answers.

**Stan Klein:** Are there any examples, nice role models of where biology has helped computational vision? Does anybody know of any good problems that biology has helped us solve?

**A.B. Bonds:** Lateral inhibition, the Laplacian of Gaussians kind of stuff is a good example. It seems to work well.

**Christopher Tyler:** How about David Marr's whole program for the analysis of visual patterns?

**John Robson:** That certainly wasn't based on what happens in the visual system.

**Christopher Tyler:** Inspired by what happens in the visual system.

**John Robson:** Ah, yes, it's inspired.

**A.B. Bonds:** Airplanes don't flap their wings either. (Laughter from the Floor) There's a certain acknowledgement of the fact that a bird's wing has got an airfoil. We're not trying to do a slavish neuron by neuron model, but at the same time, there might be fundamental principles of organization that could be helpful. What my problem is in understanding why we're here, and what

we can do to contribute to the improvement of computational models is that I have not a clue as to what the problems are in the computational realm. I don't read that literature, and I don't know what people are struggling with. So in a sense this meeting has been called to bring expertise to bear on the problems of computational vision, from the perspective of the physiologists and the psychophysicists. What kinds of problems are we dealing with here?

**Teri Lawton:** One of the problems is object segmentation. Another is automated navigation.

**A.B. Bonds:** No, automated navigation is not a problem. Automated navigation is a design issue. Within automated navigation you can parse it down to specific kinds of problems. Segmentation is a problem. O.K. But that's too general an area.

**Malcolm Young:** There were another few hands when asked who was a modeler. So, anyone else care to say what the problems are?

**Stan Klein:** Stereo is an obvious one also.

**A.B. Bonds:** In what sense is stereo a problem? Stereo is a thing. What aspect of stereo is a problem?

**Christopher Tyler:** The correspondance problem.

**Peter Lennie:** Yes, that's an interesting problem for us, but not an interesting problem for computer scientists. It's a demonstration project for computer scientists.

**Jeff Teeters:** We're trying to build machines that do that. Some people are working on stereopsis and trying to solve it.

**A.B. Bonds:** Some do, Ballard does, for example.

**Peter Lennie:** So why don't we just use range finders. These are simple things that cameras use that work very well.

**Stan Klein:** Or they sometimes use three cameras, they cheat.

**John Robson:** Yes, and that's the point. (Agreement from others on the Floor)

**Peter Lennie:** No, the point is it's an easier solution than solving the problem of stereopsis.

**A.B. Bonds:** So then why don't we have 3 eyes?

**Teri Lawton:** Range finders are very limited. The techniques that they use are very limited in their range and accuracy, since occlusion prevents measurement of distant objects.

**Ralph Siegel:** Maybe it reduces down to the brain being a bag of tricks, as Rama would say. The idea that there must be an elegant solution and that the brain is something really deep is perhaps wrong. Maybe, it is just a bunch of dirty tricks, and everything just got thrown together.

**Malcolm Young:** Perhaps it's that psychophysics tends to be done in a rather piecemeal fashion. I'm not sure that other neurobiologists think that it's a bag of tricks.

**A.B. Bonds:** Actually, yes.

**John Robson:** I'm sorry, what's the other alternatives?

**Malcolm Young:** The brain may be making some elegant model solutions.

**A. B. Bonds:** If there was an elegant solution, there wouldn't be 35 areas and 300 pathways, or whatever.

**John Robson:** The brain's clearly a mess.

**A.B. Bonds:** Yes, it's a mess.

**John Robson:** Maybe, one of the most interesting things, of course, is to know what any piece of apparatus can actually do. The sheer demonstration that you can do stereo is significant. As to how we do it, this seems to be difficult to understand.

**Malcolm Young:** It does seem to me that it's not obvious from neurobiology what are each of the important computations needed to build a robust model. You said that we don't know everything, but we can make some inferences based on numbers. It seems to me that you can also make other inferences to bring out the fact that the scale of the visual system is vastly greater than the scale of the models we can implement. You have to be sure that the kinds of principles that you extract from neurobiology are scaled down to be useful for modeling.

**John Robson:** It's interesting that you should say that, because one of the things that has happened over my lifetime is that computational systems have gone from being much less powerful than the visual system, to actually being much more powerful, without the problem having to automatically solve itself. I can take a particular example. When we first started thinking about any kind of pattern recognition machines, the idea that we might have 50,000 pixel images seemed so beyond everything, that people said we might be able to manage 32x32 images. We've got a very large ground to cover. How are we going to manage that? Now anybody can have 2000x2000 pixel images in their visual system. Certainly, the human visual system does not have more than 1000x1000 pixel images, and artificial systems don't do nearly as well as our visual system, even though computing power has gone up enormously.

**Christopher Tyler:** That's exactly why it seems as though biological insights could be gained. With the vast number of people working on computational vision, and with all this computing power at their disposal, they've still been unable to come up with a solution.

**Peter Lennie:** To what problems?

**Christopher Tyler:** The problem of recognizing that there's a component on a conveyor belt that needs to be picked up. I am referring to industrial robots.

**Peter Lennie:** Surely the whole thing is that for the system to be robust, it must be very adaptive. So if you say let's really define one problem, then you can easily build some sort of computational system that will solve just that problem. But, then it's useless for anything else. Our visual system just isn't that specialized. We're very good at doing many different visual tasks.

**Christopher Tyler:** What do you mean by solving one problem? It could be one problem under a variety of illumination conditions.

**Peter Lennie:** That's a major problem that we have no trouble with. But computationally that's very long. To reduce it down to such an extent as to measure the length of this hall. A computer system can do that far better than we can, but it's useless at anything else. So then this idea of trying to define exactly what a problem for computational aspects are is in a sense pointless, because then you're getting away from the robustness, which the visual system seems to have.

**Christopher Tyler:** You can define a task as being able to recognize some subtle differences in components under a variety of illumination conditions. That means it's got to be extremely accurate. That's a fairly benign problem, but it requires robustness.

**Lucia Vaina:** I think that one of the characteristics of most computer vision models is their very precise measurements. It seems that biological vision doesn't need this precision. Perhaps many unprecise measurements can somehow be pooled together to get the underlying signal and at the same time be adaptive. It's very interesting to see how we can do a sloppy job in a way, but still can recover higher up, and do the task higher up than the computer vision modelers, at least the ones I'm familiar with, who lack precision when trying to do a high level task.

**John Robson:** Sounds like bad engineering. It doesn't seem to me to be a serious problem.

**Lucia Vaina:** It's gone on for 20 years.

**Teri Lawton:** Bad engineering is what happened at JPL a few years ago. Right before an important computer vision-manipulator demonstration, a janitor came in and slightly moved the apparatus. Since everything had been precisely calibrated to the position of an external source, the demo didn't work. I think that computer vision systems have real problems, and must move beyond engineering techniques. They don't have the robustness as does biological vision, by using the redundancy conveyed by different object attributes.

**John Robson:** I'm sorry, but I don't think I really agree. I think that it's simply bad engineering.

**Malcolm Young:** I disagree. You say that some computational vision systems are as powerful now as the visual system is computationally.

**John Robson:** I don't know that I've recently seen any computations of the power of the brain. But it's actually quite a lot. That's not really very powerful.

**Jeff Teeters:** Are you saying the brain's not really powerful?

**John Robson:** Yes.

**Jeff Teeters:** I don't think that's true. There's 10 to the eleventh neurons. Just based on sheer numbers, the brain's very powerful.

**John Robson:** But, they're incredibly slow. Ten to the eleventh or 10 to the ninth used to be quite a big number, but now it's possible to do that many operations in a second using current computers.

**Malcolm Young:** The very largest neural nets that I know about have about 10 to the ninth neurons. Each neuron makes say 10,000 synapses with other neurons. I mean that's an awfully big number.

**A.B. Bonds and Christopher Tyler:** Not 10,000. There's 50,000 synaptic connections.

**Ralph Siegel:** Edelman's got models with an even bigger number of connections. (Laughter from the Floor) But, to make a reasonable point. We're all in vision, I think. The visual system doesn't exist in isolation. The visual system calibrates itself with the outside world and to the somatosensory and motor systems. Maybe one of the simple things we're missing is that one cannot look at primary visual cortex or subcortical areas in isolation.

**Malcolm Young:** So interactions with other systems are important. But what, if anything, do the very precise studies, that we heard this morning, about striate cortex (V1) tell us about how we should build computational visual systems?

**Voice from the Floor:** Nothing.

**Malcolm Young:** That's not possible.

**Jeff Teeters:** Some people have been using the properties of V1, Jitendra Malik with Pietro Perone, for example. The idea of his orientation scaled filters, different orientations and different scales, is a good model for texture segregation.

**A.B. Bonds:** But that's not how V1 works.

**Jeff Teeters:** It's probably not exactly how V1 works. But I think it models a basic principle that you have different cells with different sensitivities. It's the same way that airplanes don't flap their wings.

**A.B. Bonds:** There are a couple of ideas, that are in a sense parlor tricks, that can be gained from the studies about V1 this morning, like contrast gain control, which I haven't seen in any computational models yet. And it's certainly something that handles the fluctuations in contrast. If you put a decent retina on the front end, which you can go get now, (They make decent retinas that do light adaptation now), then you can get some contrast gain control. That will solve a lot of the little problems. That's not a fundamental quality of how cortex works. It's a necessity. But it's pretty obvious, if you just think about the problem.

Let me give you one other brief parallel. We've got some guys who are working on computer vision for the purposes of developing a device to feed people who are paralyzed, and they have to hunt faces, and that sort of thing. For years they were working with a system that basically had evenly spaced pixels. They had a million evenly spaced pixels. They were saying the computational load on this is killing us. I walked in and said why don't you make it tiny in the middle and fat on the outside. They hadn't thought about it. Those kinds of ideas can help remarkably. But it's just bad engineering on their part that they hadn't thought of it. I think that some of these principles are usable, but I think that the detailed kind of stuff that we're embroiled in here is not necessarily going to transfer. It's there because that's the way the brain is built, but not necessarily because that's the best way to do things.

**Russell De Valois:** When you say it's bad engineering, in fact, these are things that took an awfully long time for vision to discover. And people who don't take advantage of what we've learned over the past 100 years are going to have to go through the whole process again. It's obvious that light adaptation is important. It took a long time to realize that light adaptation is important. It's not immediately obvious to someone who doesn't pay attention to biological vision. I think this is exactly the kind of contribution that we can make, in pointing out some of the problems that we've discovered, in trying to figure out how the visual system operates in general.

**Malcolm Young:** Do you think you understand the mechanisms used by visual system?

**Russell De Valois:** Some of them. We certainly know what the problems are that have to be solved. And if one doesn't solve them, then you're going to fall on your face.

**Peter Lennie:** I think that's probably true.

**A.B. Bonds:** The thing is that the way the retina operates does not lend itself to using silicon photodiodes. I think that the rules that are being used are very useful. I don't think that the particular mechanisms by which it does it is particularly useful.

**Teri Lawton:** I don't think that you can understand the rules, until you understand the mechanisms being used to implement these rules.

**A.B. Bonds.** Oh No.

**Peter Lennie:** Actually it's the other way around.

**Malcolm Young:** David Marr would say it the other way. Do you think the algorithms and the implementation are separate?

**John Robson:** I think they're totally divorced. I see no reason at all to be interested in the implementation of the algorithms, I mean the squishy world, if that's what you want to call it. Because presumably our physical visual system is optimized using particular kinds of components. You can't have an optimal solution to something in a way separate from the implementation that you're forced to use. I mean engineers simply don't have the same system to work with. The algorithms might be interesting, but even that's questionable.

**Christopher Tyler:** Is that what you meant by mechanisms, Teri?

**Teri Lawton:** What I meant is that the algorithms are the outputs of different mechanisms. If we understand how we implement the algorithms to do simple things like lateral inhibition with center-surround mechanisms, then you can extract these algorithms and implement them efficiently in a computational vision system. The way I look at it is that we have a hierarchical system, so it's important to understand what's going on at the early stages, and the various intermediate stages, as well. If the mechanism consists of several levels of processing, as do those in the visual system, then much simpler algorithms will be needed at each level.

**Peter Lennie:** Don't you want to know what these algorithms are for? Surely, you want to know what you want the system to do, before you want to worry about the details of how it does it.

**Teri Lawton:** Right.

**Peter Lennie:** The problem for us is that we don't know what we want the system to do. We can technically structure the description of what is there. But beyond that you don't know what you want the system to do.

**Teri Lawton:** For example, you know that you want it to do object segmentation.

**Peter Lennie:** But which objects?

**Teri Lawton:** Any objects that you would want to see in the scene.

**Peter Lennie:** Which scene?

**Teri Lawton:** Any scene. You have to be able to generalize.

**Peter Lennie:** But we don't know that we can deal with any scene. We deal with the scenes we have, but we don't know if we want to generalize. We don't know if we want to invest time making a machine that can deal with any scene, since you usually only want the machine to deal with a few scenes.

**Teri Lawton:** I don't think so.

**Christopher Tyler:** What about the scenes in the world we live in? What scenes are you going to limit it to?

**Peter Lennie:** Until you know how expensive it's going to be to make a machine to deal with any scene, you don't know whether the right thing to do is to have five hundred different machines, each designed for a particular task, or one that can deal with many different scenes.

**Teri Lawton:** Economically it's not feasible to develop all these different machines, one for each type of scene that you're interested in.

**Peter Lennie:** How do you know?

**A.B. Bonds:** Quite the contrary, there's hundreds of industrial vision systems used in manufacturing that are very specialized to do a simple, specific function.

**Teri Lawton:** That's true.

**Christopher Tyler:** What you need is a vacuum cleaner to clean the house, and that's the great general task.

**Peter Lennie:** No it's not. It's a very specific task in comparison to what we do visually. You can construct a pretty good description of what you have to do. You need a very small part of your visual system to do the task.

**Malcolm Young:** One says yes, and one says no.

**Christopher Tyler:** You have to recognize furniture, and the location of furniture relative to the layout of the house. Things can be constantly changing in the house. The problem is a fairly general one, because it's got to work in any house, and houses are full of most of the objects that we encounter in the world.

**Peter Lennie:** It's certainly not the right way to design machines to vacuum houses, that is, if you want to vacuum any house.

**A.B. Bonds:** You can put in a program of your house.

**Teri Lawton:** But how are you going to do that for any house?

**Christopher Tyler:** It's got to work on any house. Otherwise, you wouldn't be taken seriously. Are there computer vision systems that address these issues at some level?

**Lucia Valna:** I think that within the modeling field there is not very good data, as you know. There are people who are using analog VLSI (very large scale integration) to do the more traditional established types of approaches. The demands on the system among the people building artificial retinas are very steep. Analog VLSI is extrodinarily fast, cheap, and you can do reputable things with it, and you can vary combinations with it. If you are using regular computers, then they are much larger, slower and much more computationally intensive, but very likey to be able to solve the more complex problems. I dont' know the details of the algorithms they use. But I do know there is not a modeler's group who is particularly happy with the way the artificial system works to solve vision.

**Christopher Tyler:** How about silicon retinas?

**Lucia Vaina:** Silicon retinas are something which is one of the concerns of the analog VLSI people. There are several people working on this. Carver Mead is one, and the MIT group is another. The MIT group is doing a project using Markov random fields, that I think is very interesting, but it's not yet at the stage of using algorithms that solve motion correspondence, stereo correspondence, and so forth. They are not yet put together.

**Christopher Tyler:** There is no inherent difficulty about doing that?

**Lucia Vaina:** I don't know.

**Jeff Teeters:** I don't think there is, because a lot of the computations in the retina are fairly local, and they are quite amenable to making shifts. If you get into things that have to be done over longer ranges, correspondence, for example, where you've got two inputs coming from two different shifts, then you've got problems.

**Christopher Tyler:** Can you use that type of retina as an input into a stereo system?

**Jeff Teeters:** Yes, but then you're not using just analog VLSI.

**Christopher Tyler:** Hybrid is one of the messages of the brain.

**Lucia Vaina:** Well yes, they have not yet done that. They have tried to design a system to work as a biological retina, not one that is hybrid. They want to simulate biology by using analog VLSI to replace some functions. They haven't tried to ground it experimentally or otherwise. What happens to determine the best function is a very important question experimentally, but one that will be asked once they have it working. They have an extremely narrow focus given that they want to simulate biology. I haven't heard of anyone that asks questions that would be of interest to those studying biological motion vision. The gap is extraordinarily large, and rightly so, until they use hybrid vision.

**Malcolm Young:** Whatever John Robson says, (Laughter from the Floor) the primate visual system does seem to have a logical design. What do you think? This is an attempt to bring in the neurosystems neuroscientists studying the alert macaque monkey. There is a question that we want to make, whether we use analog VLSI or simulate it on a computer, kind of a monkey see, monkey do version of the visual system. I would like to address the first of the interesting questions Teri prepared: What does the data from neural mechanisms indicate about computational approaches that need to be implemented to account for complex motion and form analyses in natural scenes? Do we actually know that? What do we know?

**Ralph Siegel:** What do we know in the striate cortex? Do we know the methods that are used for orientation tuning?

**A.B. Bonds:** No. We have ideas, but we haven't got it down.

**Russell De Valois:** Why does one need to know the mechanism? It seems to me that a critical thing is knowing that it has to happen.

**Ralph Siegel:** Looking for analogous types of functionality led John Maunsell to the obvious experiment of looking at attentional effects on the output response of cells in MT, where the direction of motion is being monitored. We could get a list of all the high level components, above striate cortex, that are important.

**Christopher Tyler:** One of the interesting things in that list are the attentional effects. I think that the first time I've heard of anyone putting that in their computational model is in Teri's talk. I'd like to know: How do you put attention in your model?

**Teri Lawton:** Well, it's an event-based model that analyzes one object at a time. In that way, the model attends to the object, using local processing. Then algorithms, that I discovered from psychophysical studies with partially sighted observers, are used to enhance object segmentation.

**Christopher Tyler:** How does your model decide which object to attend to?

**Teri Lawton:** Right now the model starts scanning the scene, looking for objects closest to the camera, and then scans the image a line at a time, until the most distant objects are segmented. This is not the way people would scan the scene. They would want to be able to look anywhere in the scene, and start scanning from that point outwards.

**Christopher Tyler:** What's close to you is in what coordinate space?

**Teri Lawton:** When you take a picture, what's closest to you is in the bottom of the scene, and when you scan up, it usually corresponds to what's further away.

**Christopher Tyler:** You haven't implemented a lot of biological attentional strategies yet, like ones that expand and contract the field of attention?

**Teri Lawton:** Currently, the field of attention is expanded or contracted at a rudimentary level, since based on the output of boundary detectors, objects are constructed that have different sizes, in terms of their height and width. I'm setting up the core of the model right now. I think it's important to envision all the different functions and levels of processing that you want to incorporate in the model. All of the elements at the front end should be implemented, before using complex attentional strategies. Since the model uses efficient event-based scanning, a feature not seen very often in other computational models that tend to use time-consuming exhaustive scanning, it's set up to incorporate different attentional strategies. The model is only a good beginning. We need a lot of work in that area. Unfortunately, people do not appear to be developing models in this direction, at least that I'm aware of. I think it's like Russell said, we've solved certain problems, so we can take advantage of what we know about our visual system. I don't think that we need to understand mechanisms like orientation selectivity, and why we have it. However, we do know orientation selectivity exists, and should be included in our computational models. Since most other models are pixel-based, they can't incorporate oriented filters to analyze the scene. We need to develop object-based models using filters that integrate information over space.

**John Robson:** That raises an interesting question. You actually discuss using some particular kind of filter, because it seemed to be a biological one. What we know about the biology is that it's a good deal more complicated than that. We know that you don't have just one kind of filter. We know that you have lots of kinds of filters. Do you think that because you have orientation selective devices of various kinds of bandwidths, that any device of this sort necessarily ought to incorporate orientation selectivity somewhere inside it?

**Teri Lawton:** Are you talking about using even- and odd-symmetric orientation tuned filters, like those that I've included at the front end of my computational vision system, to extract boundary and texture information?

**John Robson:** Yes, that's another example. We know for the sake of argument, that the visual system doesn't actually have even- and odd-symmetric filters.

**Teri Lawton:** At the front end or anywhere?

**John Robson:** Anywhere, to my knowledge.

**Teri Lawton:** You don't think that we do?

**John Robson:** Izumi could answer this better than I. I think the current studies indicate that is not what we actually have. Even- and odd- symmetric cells were an idealized earlier notion. But the best information we have now is that is not correct.

**Izumi Ohzawa:** It may not be exactly even or odd. The basic idea is that you have to have this even- and odd- symmetry in the visual system.

**John Robson:** What we actually have are not even- and odd- symmetric ones, even if we have ones in quadrature pairs. We don't know, I think, at the moment, whether there is something particularly desirable about what we actually have, or whether it simply doesn't matter whether it's that or the actual even and odd ones. I think that some people are promoting the view that if you look in the visual system and find not actually even and odd symmetric cells, then this is what you should incorporate. We may not know why it's desirable to do that, but we have to do it that way, because that's how the visual system does it.

**Russell De Valois:** I certainly didn't say that one has to incorporate the different operations we know that the visual system does. If one tries to build something, and fails, which in fact is what's happened in computational vision, where people thought that they could solve the whole problem, then a sensible approach is to take some system, and to incorporate features of that system, even if you don't know exactly how it's put together, or perhaps not even its function.

**Peter Lennie:** We don't know how all the neurons in the brain work, except in some obscure way. We know that the neurons do something. We know what happens locally when we make a measurement, but we don't know what the result of that operation is after the measurement is made.

**Russell De Valois:** The alternative is not doing anything. That's the point. If one understood how the whole visual system works, and you could design a device without paying any attention to the brain, fine. There's certainly no harm in incorporating the operations we know that the visual system is doing. The brain is too complicated to understand how it all works.

**John Robson:** You have to know how to incorporate orientation detectors into whatever you're building. You can't just have them.

**Christopher Tyler:** It has to be a bootstrap procedure, where you incorporate these detectors, and work with them to see if this operation does any good. If it does you keep it. I think the brain can be a source of inspiration, rather than providing the definitive answers.

**Peter Lennie:** I think that only in a limited sense what Russell says is true, that if you don't know what to do, then what we know about the brain can help a lot.

**A.B. Bonds:** Let me draw a brief analogy. It sounds to me like we're talking about how the brain that is filled with different cells tuned to different orientations and spatial frequencies works. A computer is built out of integrated circuits. So what I'm going to do is go down to the integrated circuit store and buy a whole bucket of integrated circuits and wire them together just like this other computer is wired together, and then I'll have a computer that works, except I don't know what's inside the box. I think that's part of our problem. We characterize, I think in a sense we're anthropomorphizing the cells into what we think they ought to be. We think they ought to be

orientation detectors and Gabor filters, because that's what we invented to measure them with, and in fact, they may not be those things. They may be something else constrained by the developmental requirements of the brain. They're the closest compromise that we can come up with, but we haven't necessarily found the best tools for making these measurements. What we're trying to do is build something out of bricks. But we don't know how big the bricks are, or how long or wide they ought to be, or how they fit together. But we're going to make something out of bricks, by golly, because things are built out of bricks. I think that's really a bankrupt approach.

**Peter Lennie:** Rather than starting off building it out of mortar or straw. (Laughter from the Floor)

**A.B. Bonds:** I'm an engineer. I'd sit down and say what's the problem and do a top down design, instead of a bottom up design.

**Russell De Valois:** That's what people have tried and failed at. That's the point. If it were a simple system that one could design without paying attention to the brain, then obviously, that would be the way to do it. The point is that people have failed. It's just too complicated a problem.

**A.B. Bonds:** There are hundreds of industrial systems that work extremely well for very narrow, dedicated tasks. The question is: Is it really necessary and desirable to come up with or make one generalizable, robust system?

**Peter Lennie:** Could they take all of those hundred systems and put them together to make one system that would work?

**A.B. Bonds:** No, it wouldn't work.

**Lucia Vaina:** The industrial systems don't care about biology. They only want it to work. I think that in computational vision, people are trying to take into account biology. I think that the question that needs to be answered is the following. In 1980 when David Marr died, he and his group left a computational vision theory that inspired many others in computer vision. He included stereo and was looking around for other theories. Perhaps if we looked into these theories, the problems, solutions, approaches, and the experiments that were designed to test out these ideas, at what computational vision did 15 years ago and what it does now, then we would have more success. I think it's a different question than what computer vision does or does not do.

**Peter Lennie:** Computational vision is doing exactly what A.B. was talking about. The problem of constructing a general representation has become much too hard. All the computational people I know abandoned that approach 15 years ago, because it's intractable. There are no global solutions to well defined problems.

**Lucia Vaina:** Except there were two specific algorithms in 1980 for stereo proposed by Marr and Poggio. What is the evidence for and against? My feeling is that this began in 1966 and wasn't abandoned until 1980.

**Christopher Tyler:** I spent the last 15 years getting up to speed on the computational techniques and the issues. So, I am just at the threshold of being able to address some of those questions. It's clear that some of the things that came out of Marr's model have run into trouble, such as the zero-crossing concept. It's been running into difficulties as a mathematical theorem and as a model of visual processing. Morgan and Watt's work suggests that this is not actually how the visual system works.

**Lucia Vaina:** Right. So they have made the suggestion that zero-crossings are not a good thing to look at for stereo, as well.

**Christopher Tyler:** So, I think that the status is that some of the difficulties are being brought out.

**Lucia Vaina:** Right. Now that some of these difficulties have been brought out, does anyone know what kinds of updated models have been proposed by computational vision people to take into account these difficulties?

**Peter Lennie:** The approach of trying to construct a general representation of the external world has been completely abandoned by those in computer science.

**Stan Klein:** That's a pretty global statement. (Laughter from the Floor)

**Peter Lennie:** Some of the people here may not know what's been going on. The focus for the past four or five years has been very directly on dealing with local problems with very local solutions.

**Christopher Tyler:** What about Jitendra Malik?

**Stan Klein:** Malik's representation is that there are these filters, Hubel and Wiesel type cells, and everything is interpreted in terms of these filters.

**Peter Lennie:** What's being done with it?

**Stan Klein:** It's kind of a John Robson talk. (Laughter from Floor) I'm going to speak for Malik. Other people correct me if I'm wrong. He claims that his motion flow algorithms that are based on looking at things through even-symmetric filters, which is different than how others who are trying to get motion flow fields, is a much more successful algorithm. His stereo uses slant and shape from texture gradients. By looking through the filters, which is kind of like human biology, apparently it is very successful for doing things that people who have been using other techniques are not doing as well, such as detecting textures and their segregation. Looking through these filters is a very nice representation for what the higher levels should be looking at.

**A.B. Bonds:** You are saying what the higher levels should be looking at. But we don't know that. How much of this is really biological?

**Stan Klein:** I'm not sure.

**Russell De Valois:** Certainly, the inspirations for Malik's work is the recordings from cortex. Whether it's correct or not, that's in fact where the inspiration came from.

**Stan Klein:** I'm just reacting to Peter's point that the computer vision people have abandoned any general representation.

**Peter Lennie:** I believe that to be generally true, based on the ones I've talked to.

**Stan Klein:** Well, Malik might be an exception, but I think he's going a very good direction, because he's been very successful.

**Christopher Tyler:** How about Beau Watson or Roger Watt and image compression?

**Lucia Vaina:** Do you really think people have abandoned a general representation in their computer vision models?

**Peter Lennie:** The approach that was characteristic of the work, say particularly in David Marr's book, was to construct a general representation upon which you could draw to make perceptual decisions to drive actions. And that's not what people are largely doing right now. It's too complicated and intractable a problem to get a generalizable representation.

**Michael Oram:** Biology suggests that a generalized object concept is not correct either.

**Lucia Vaina:** Actually to go back to my question: What happened to Marr's legacy in terms of the object centers that we've been talking about? What's happened, at least in some groups, is that there are biologists like Dave Gray who have been suggesting in the past few years that there are useful tools to be used for object representation that work pretty well. Now people in computer vision tend to suggest that there are indeed some specific views from which you can generalize. You can combine several specific viewer-centered representations, so that you can interpolate between these specific views to get any other views in between. They find that this is a very good combination, a very efficient one to achieve object recognition. So, this has been done in monkeys by Logothetis, in psychophysics by Tommy Poggio and by Shimon Ullman. It's being done by a group of people who brought themselves together to pool biology with psychophysics to develop good models. That's one way it can be done. I would like to note that this idea of multiple viewer-centered representations being combined into a generalized representation is not new, and comes from research done in the seventies, that I remember reading about.

**Christopher Tyler:** In terms of people who have global models, how about Gerald Edelman, Terry Sejnowski, Stephen Grossberg, Ennio Mingolla, Jitendra Malik, and Marvin Minsky. There's a number of people who are trying to do big chunks of the problem. You may think that their solutions are not great. But it's not that they've abandoned the effort.

**Anne Treisman:** There are learning aspects, too. Letting neural networks develop their own learning mechanisms by giving feedback.

**Christopher Tyler:** That in itself is a biological mechanism, an evolutionary strategy.

**Anne Treisman:** You don't understand the models when they emerge. These hidden units, nobody knows what they're doing.

**Christopher Tyler:** That's right. What's your idea about evolutionary techniques in your model, Teri. Are you using that, or is it something for the future?

**Teri Lawton:** What do you mean by evolutionary techniques?

**Christopher Tyler:** Training.

**Teri Lawton:** I've implemented the code that uses the redundancy of object attributes to match the same object in subsequent scenes. Right now I'm improving the object segmentation algorithms, before proceeding on to test the training algorithms, that are already implemented. I have a sequence of 12 views of the same scene, where objects differ only in their translational movement from one scene to the next. Therefore, using motion parallax to determine the depth of each object, updating the depth estimate with each new scene, is a very straightforward task. In addition to updating the depth estimate with each scene, the effects of occlusion are removed when possible. Matching is based on the assumption that the top of the object is more likely to be uncovered than the base of the object, when occluding objects are present. With each subsequent scene more of the object is uncovered. Each object consists of a dataset of attributes, such as height, width, mean luminance, depth, and so on. Each time the same object attributes are found, within some range of values, then the probability that these attributes are accurate is increased, this probability being a component of the object dataset. If you get the same object width and height in at

least two scenes, then there's a high probability that you've segmented the object accurately. Bob Snowden and Ol Braddick have shown that at most 4 to 5 scenes are needed for observers to optimize direction selectivity based on motion parallax. Therefore, it seems that the most you would need to construct an accurate 3-D object map is 4 to 5 scenes. This type of training or learning is implemented in the computer vision system I developed to improve the robustness of constructing 3-D object maps. I think that it works in a manner similar to the way people learn.

**Jenny Lund:** People spend from 3 to 5 years, from birth onwards perfecting these circuits. I wonder if any of the computer scientists have the patience to let their neural networks run for such a long period of time.

**Christopher Tyler:** With no chance that it would crash in that amount of time. (Laughter from the Floor)

**Jenny Lund:** Maybe, this is the secret among human visual systems. They have taken a long time perfecting themselves in natural scenes, and responding to the natural scenes in various ways. It is really a system that was built with a purpose.

**Christopher Tyler:** I think that a lot of the visual processing is done in eight months, or a year.

**Jenny Lund:** By the time children are learning to walk or crawl. It may turn out that they are more sophisticated at things like face recognition, which I have never mastered. (Laughter from the Floor) Things like face recognition and reading take a long time to develop.

**Peter Lennie:** It's not clear that reading is an important task of the visual system. We need to know how something is going to affect the organism.

**Michael Oram:** We are not subjected to reading as much as other patterns. Therefore, we are exposed to other patterns at a much younger age.

**A.B. Bonds:** There's a problem with hidden units. We don't know what they're doing. What strategies they're using, like the wait state analysis.

**Peter Lennie:** There are two cases that are used as demonstrations. One by Zipser and Anderson, and one by Leaky and Sejnowski. In both cases, if you actually think about what properties you might expect, then once you have a system that will do the job at hand, the hidden units don't have properties that are altogether mysterious. It's as if these properties aren't completely untouchable, without knowing about them.

**A.B. Bonds:** That's when you know what all the hidden units are doing. The wait state analysis that's used in those examples are based on a knowledge of the population of hidden units, what you're feeding all the hidden units from all the input units, and what all the hidden units are doing. If you just look at one hidden unit in isolation, then it's very hard to fathom what the system's all about.

**Michael Oram:** When you stick an electrode into the brain, how easy is it to define what those units do? It seems that most of the recordings from the macaque monkey brain have a very explicit code as to what the cell is responding to. Is this the case with the hidden units or are they very wishy washy, seeming to respond to a lot? If that's the case, then it seems that the macaque brain is very different, having a very explicit code of some nature which you can easily tap into to.

**Jenny Lund:** Is it so easy to get the right description of what the cell's responding to?

**A.B. Bonds:** That's a very good point.

**Michael Oram:** You probably can't know everything, but you can say it responds to this stimulus much more strongly, than it responds to this stimulus. Is this the case for the hidden units?

**A.B. Bonds:** Sometimes it is. But what scares me is that when I record from a cortical unit, I know what I can drive it with, but that doesn't tell me anything about what that unit really wants.

**Christopher Tyler:** I have a question about the hidden unit analysis. My understanding is that Tom Albright and Terry Sejnowski share these mysterious things that they hide from the outside world. (Laughter from the Floor) Unit properties that physiologists have thought are computational. Then they'll discuss it over lunch, and they'll discover that these properties aren't as mysterious as they were supposed to be, especially when you find out that the hidden units in the computations have similar properties to the cells.

**A.B. Bonds:** There are some examples of that. The shape from shading algorithms, for example, of Sejnowski and Leaky. They came up with orientation selectivity which was a necessity in order to do the shape from shading task. That gets to another philosophical view of things. What that tells you, from my perspective, is that if you're going to do shape from shading, you have to have orientation selectivity. So, the neural networks are telling you the things that are required to solve the problem. Maybe the brain will be telling you the same kinds of things, the things that are required to solve the problem. This is a problem defined architecture. We are working on solving a very specific problem. Therefore, we can go looking for solutions to that very specific problem. You will not find neural networks that will do this kind of global vision task. They're not going to tell you, therefore, the magic behind doing the global vision task. They will tell you specific things very nicely, but they won't tell you how this whole multipurpose system works.

**Christopher Tyler:** Maybe there's building bricks. Maybe you need to solve shape from shading, then solve segmentation, then get an object-based map of a 3-D scene.

**A.B. Bonds:** Is that why we have 35 areas?

**Christopher Tyler:** I think it could well be. Absolutely. Why not?

**Michael Oram:** Do you think that there's a way, some advantage of having a couple of stages and then vision is solved. If it's that simple, then why haven't we done it?

**Anne Treisman:** Why would face recognition have its own area, if the brain wasn't divided into tasks?

**Malcolm Young:** In monkey cortex, there is not one area devoted to face recognition. In people, lesion studies indicate that there is a face recognition area. In monkeys if you chop out the anterior part of the cingulate cortex or the superior polysensory area, you still get face recognition. What you see is that other aspects of vision get their attention. You still leave some face recognition cells. There is a distribution around the brain where you see face cells. There are perhaps a dozen, at least a half dozen areas that have face cells.

**Jenny Lund:** Perhaps, monkeys need more than the face of individuals for recognition, like smell, or seeing the whole body, for example.

**Michael Oram:** It's certainly a problem when you're asking the monkey to recognize a face. It's very hard. You can't ask the monkey to do the task. You've got to teach it, and the variable may not be the one you're expecting it to be using. It may be that the monkey is looking for one eye. If the monkey sees one eye, then he says it's a face. In trying to teach the monkey a task, they will use

what ever method that they can to solve the task. If you try and change the task, they'll just go for some other method that solves the task. The interpretation of the behavioral studies are very hard, because you're never really sure what they're actually doing.

**Jeff Teeters:** One thing that hasn't been mentioned yet, and that I'd like to hear people's opinions on, is how does high level vision relate to lower level processing? I think that a lot of the tasks that are done in vision, even segmentation, for example, are human's color blind or anomalous, are some of the roadblocks that prevented people from developing general purpose vision systems. Do people think that's the case or can we get by with some heuristics to implement these tasks?

**Malcolm Young:** If there are all these feedback pathways from higher visual areas to lower ones, then why hasn't anyone ever recorded a cortico-cortical back projection? We can see back projections from V1 to LGN, but not from V2 to V1.

**Leslie Ungerleider:** There are minimal effects. We cannot see the activation from cortico-cortical feedback, because the processing has to be going on at the lower level area, before the higher level area is activated to generate feedback. What we need to do is look at the processing that is going at lower levels after higher levels have been deactivated.

**Ralph Siegel:** One of the very powerful things is a moving texture that cannot be seen at a lower level area, in contrast to the direction selectivity of a moving grating that is seen at lower levels.

**Malcolm Young:** It has been shown that if you can measure cross-correlations among synaptic branches, then you can demonstrate that there are lots of routes that lead to back projections. For that particular effect you don't have to have cortico-cortical feedback.

**Christopher Tyler:** But what about any other back propagations?

**Malcolm Young:** Back projections relate to the field surface that has to do with topical dendrites. Presumably, they could run down through the dorsal pathway and come back up again. It does seem odd that these very common projections seem to be so difficult to record from.

**Ralph Siegel:** John, how do you get somatosensory input from V4, if it weren't for back projections?

**John Maunsell:** We thought we'd see back projections. Frankly, I was disappointed. We went intracellularly, and looked at saturations, modulations, changes in the response properties, and could find some changes in V1. There is some physiological evidence for back projections. But we didn't find anything like we expected. We had to ask whether the effects of the processing at higher levels on the processing at lower levels could be measured with the techniques that we're using. The best examples we have come up with are from tasks that require higher levels of processing, where attention is needed to fire the cell, such as found for cells in the parietal lobe, as described by Carol Colby. However, we need to be able to measure the effects of back propagations on tasks determined at lower levels of processing also.

**Malcolm Young:** Also, they're attention effects that you find in the amygdala projections. Just because there are modulations that clearly aren't retinal doesn't mean that they are driven by are cortico-cortical back projections.

**Christopher Tyler:** Then how could you demonstrate the cortico-cortical back projection?

**Malcolm Young:** From cross-correlations.

**Christopher Tyler:** But not many people have even looked at cross-correlations.

**Ralph Siegal:** There are oscillations that have been found which would give you cross-correlations between areas V2 and V1.

**A.B. Bonds:** It doesn't necessarily mean that V2 is going back to V1. They are probably driven from common input. There wasn't a high enough temporal resolution to show that it was definitely a back propagation.

**Ralph Siegel:** I think that people are concentrating on measuring forward projections, before trying to measure back propagations. I don't think that means there is no evidence for back projections.

**Malcolm Young:** All I'm saying is that there are no direct demonstrations of top down interactions that are definitely cortico-cortical back projections.

**John Robson:** On a short time scale, on a signal processing time scale, rather than a much longer time scale. There is a whole question of how the system calibrates itself, where I don't think you'd expect to see things on the type of time scale we're discussing anyway.

**Michael Oram:** I think it all comes back to what particular aspects of back projections we're thinking about. The question is to what level do you think that we could computationally be doing something.

**Jeff Teeters:** I think it is that the presumption in general that higher level knowledge is essential for a lot of the perceptions that occur.

**Peter Lennie:** You can address the question of segmentation. There's a few celebrated examples where you clearly have to know what's what to understand things. The dalmatian dog, and various special figure-ground combinations. But, most of the time segmentation is effortless. When you are talking about segmentations that are not effortless, then feedback is probably needed. The acid test is can you do the segmentation quickly with no experience.

**Michael Oram:** There is also the level of the experience in setting up the feedforward projections. The back projections may be important in setting these up. Once you have experienced it, then you have set up the weights that are used subsequently. Then you don't have a purely feed forward system, you use the back projections to set up the weights for that particular task.

**John Robson:** Then that's a very long timescale.

**Michael Oram:** Yes. It's a rather dynamic system.

**Ralph Siegel:** Edelman's recent modeling indicates that it would be important to look at reentry that leads to modifications in the output response, and is needed to make the brain work.

**Malcolm Young:** There is one last idea. It's not the brain that sees things, it's just neurons connected in fancy ways.

**Christopher Tyler:** What about the glia? (Laughter from the Floor)

**Malcolm Young:** There are only two aspects that all we've abstracted from neurobiology. One is that we know some aspects of the biophysical processing. Another is that we know that there are lots of aspects to the connectivity. And that's the only other thing that we've derived from the

neurobiology for orientation detectors, gain control, or face cells, or whatever. So perhaps we ought to pay attention to some aspects of the differential connectivity.

**Russell De Valois:** I don't understand the premise. It seems to me to be just the reverse. Computational vision does pay attention to the connectivity, not to the biophysics.

**Malcolm Young:** For example, you always have full connectivity between all the elements. Everything is connected to everything else.

**Russell De Valois:** What else is there?

**Malcolm Young:** What else generates these properties? Where do orientation detectors come from, if it's not from the biophysics, and, for example, interactions between local connectivity and extrinsic connectivity? What else is there?

**Christopher Tyler:** What kind of connectivity are you talking about?

**Malcolm Young:** Local connectivity that depends on the kinds of patterns being presented. These mechanisms have different functions.

**Christopher Tyler:** Magno parvo separations, that sort of thing?

**Malcolm Young:** All that you've got is the processing that is a result of the particular biophysical properties, and perhaps you've got connectivity.

**A.B. Bonds:** One of the unfortunate problems that we face, however, is that we either know what it does, but not how it's connected, or how it's connected, but not what it does. It's very challenging, incredibly difficult to mix the two, to get information on the fact that it has a certain shape and is connected to certain things. I don't know if we can learn anything.

**Jenny Lund:** Certainly it would be interesting to try and construct some simple network. Now, I don't know enough about it to know if it's ever simple to do these things, but whether people could play with these different scales of local inhibition, whether it would force connectivity that if left free would run about and choose to connect to punctate cells, and why is it that most of cortex has this structure? Does this mean that you can feed vision into almost any visual cortex? People have moved afferent inputs from different bits of the thalamus into inappropriate cortex, and it apparently comes up with oriented units. So maybe there are some general rules about cortical connectivity that give it a quality that is especially useful, in terms of network construction. Maybe it would be worth playing with some of these observed structures just to see what would happen if you fed in, say two or three different qualities. Would they sort into gradients, and would they oscillate into repeating effects? It would be nice to see if there were some interesting effects of computations. I don't know if network people have thought about these ideas.

**A.B. Bonds:** There are radial basis types of networks that certainly address center-surround organization.

**Christopher Tyler:** There was that fellow at Stanford or UCSF that was modeling ocular dominance columns, Ken Miller.

**A.B. Bonds:** There are a lot of models out there that are trying to do that sort of thing. There are at least a half-dozen ocular dominance computational models that solve simple vision problems.

**Jenny Lund:** But just to guess, I'd say that ocular dominance is perhaps just one of many possible paradigms. If you want afferents to sort into discrete separate zones from one another, that's one

property. If they want to construct gradient properties between them, that's another issue that looks interesting. So I think that there's very fertile ground for people interested in making neural nets to try some different constructs, just to see what would happen.

**Peter Lennie:** The structure of the cortex that receives visual input has changed so that it looks visual whereas the somatosensory primary cortex has a very different structure. It seems that the functional structure of the primary cortical region depends on the type of input that it has. Somatosensory seems to have in certain ways different structures than the visual. If you swap the two around, you get the change in the extrafine structure between the two areas. It seems that the structure depends on the input that it receives.

**Jenny Lund.** I think that's a particularly interesting event. The characteristics of the thalamic input, particularly, perhaps, the way they segregate, may force a structure on the postsynaptic cell groups. The interesting thing that could be played with is why  $\alpha$  (magno) and  $\beta$  (parvo) have different cell packing densities, and why that might come to be. It seems to be matched in a way to the different densities of the afferents. And then there's the barrel fields for whisker inputs in somatosensory cortex that seem to crystalize our response to the afferents coming in. So in a way your neural fields are shaped, very much by, to start with, these oscillations that structured the thalamus. So I don't think that modelers should be wary of making presumptions that specialize your neural fields to your afferents. They seem determined to connect everything to everything else. I just don't believe that's the way the nervous system works. I think that modelers should be much more creative about the patterns they analyze and structures they implement.

**Malcolm Young:** I agree. This is a good point on which to end this workshop.

## Parallel Processing in Monkey Extrastriate Cortex

John H.R. Maunsell\* and Vincent P. Ferrara†

\* Division of Neuroscience, Baylor College of Medicine, Houston, Texas 77030

+Department of Physiology, P.O. Box 0444, University of California, San Francisco, CA 94143

Extrastriate visual cortex in primates can be divided into two distinct streams of processing, which subserve different visual functions. A dorsal pathway contains areas in the parietal cortex, and is thought to be important for processing information related to spatial relationships and movements. A ventral pathway contains areas in the temporal lobe, and is more involved in tasks related to visual identification and recognition. We have examined extraretinal signals in these pathways by recording the activity of individual neurons in macaque monkeys while they perform match-to-sample tasks. In previous studies we found that neurons in the temporal pathway conveyed signals that appear to be related to memory of the orientation of a sample stimulus. Recently, we have looked for corresponding extraretinal activity in the parietal pathway. Because the parietal pathway contains many direction selective neurons, we used a direction match-to-sample task. Relatively little evidence for extraretinal signals was found in the parietal pathway during direction matching. Instead, we found such signals were prevalent in the temporal pathway, although motion processing is normally associated with the parietal pathway. The apparent involvement of the temporal pathway in matching stimulus motion is consistent with the notion that the parietal pathway is specialized for motion analysis related to visual guidance, and is not involved in all classes of motion processing.

**Jenny Lund:** I wondered if the way you presented the stimulus could have something to do with your findings. In other words did the monkey brain treat it as an object attribute, the motion on this train. If you had presented it as an object moving against the background, or something of that kind, would it have enlisted the parietal pathway?

**John Maunsell:** I think that's very likely. Of course, I don't mean to imply that there are fewer extraretinal signals in the parietal pathway. I think that it's largely dependent on the task the animals asked to do. So, if you'd asked these animals to make eye movements to remember targets. I suspect it's the inverse situation, where most of the extraretinal representations would be in the parietal pathway. I don't know how the animal approaches the task, I don't think there's any way to know that, but I think it's very likely that it's treating those different directions as objects. Any one of us doing the task would instantly recode the sample into left, right, up, or down and as stimuli came on, compare it to that. Rather than trying to create some sort of mental image of the dots flowing left, when the sample is presented in that direction. Obviously, monkeys don't have language capabilities along those lines. But it wouldn't surprise me at all if they don't have a representation that is fundamentally object-based, when you ask them to remember a particular patch of dots moving in one direction. I think that if we could get them to have to deal with the dots as a trajectory, as a prediction of where things would end up after a period of time, then it's very likely that we would see much more activation in the parietal pathway.

**Ralph Siegel:** Do you know Jim Gnadt's work showing very long attentional effects in LIP, and also Mountcastle's peripheral dimming effects. I think it's like you're suggesting, if you have the animal doing a spatial task, then you're probably going to see these effects. It makes sense from the lesion work in area 7a, as well, where you have attentional spatial deficits, form type deficits, and remembering deficits. I think that the task, as you say, is probably the critical thing.

**John Maunsell:** Yes, I think that my view of it these days is that there are still questions as to how prevalent these signals are, as you get to the very later stages. We were dealing with V4 and MT, which are relatively early on, and the effects were not very overwhelming, in terms of their

numbers or their strength. That's an open question. I really have the feeling now that there's a very rich repertoire of representations in extrastriate cortex, in terms of abstract entities and motivational state. I think it would be very valuable to have a good handle on that, and how these different representations get sorted out among these areas. It would be very nice to see whether there was as much specialization for extraretinal representations, as there was for straight sensory representations.

**Leslie Ungerleider:** I wonder if you would speculate on where you think these extraretinal signals are coming from. Are they from parietal or ventral stream areas, or do you think they derive from frontal areas?

**John Maunsell:** That's really speculation. I have no idea. I'm inclined to suspect parietal less, only because our recording up there didn't suggest that region was being activated by these sort of effects. I'd very much like to know what's going on in inferotemporal cortex, and whether that was one of the stations on the way back to V4 for getting those effects in V4. In terms of whether its predominantly prefrontal in origin, amygdala in origin, basal ganglia in origin, I really don't know. There are a lot of studies that have been done in prefrontal cortex similar to this, so I think that's a very good candidate.

**Christopher Tyler:** When you showed the matrix of 16 responses, I was expecting to see an interactive response. It seems to me that if the cell is encoding what the animal is looking for as occurring when the stimulus was presented, then you should get a big difference on that one cell when the two would intersect. It seems to me that you only showed that in one of the slides, and you didn't analyze across the population, the extent to which that interaction occurred.

**John Maunsell:** That's right.

**Christopher Tyler:** A further point would be, in the cell that responded to all four directions of motion, you might expect to see a response on the main diagonal. I think I saw suppression in that case all the way down the main diagonal. Have you analyzed that at all?

**John Maunsell:** Yes, we did. There's a lot of interactions between a cell's preference for a particular stimulus orientation, and its preference for a particular sample orientation. It's not uncommon to see interactions where the cell will be largely responsive to one of the 16 conditions. Frequently, that's a matching condition, as you saw on one of the slides. In some cases, it's a nonmatching condition, which we found surprising at first. We've come to live with it now, accepting that it has as much information as a cell that provides information about when a match occurs. We certainly see some cases where there are nonmatching conditions that give the best response overall. We never saw in any of the cells in any of the areas we've recorded from, a cell that responds best to all four of the matching conditions, and less to the remaining nonmatching conditions, or the converse. We see a cell that responds to one match, or two matches. But we never see a cell that has a complete solution to the task, and is active when the animal should release his hand.

**Ted Cohn:** If this is really a central influence acting on the cell, one would be curious to see the results of a test where you put the stimulus to be remembered outside of the field that you're testing in, or develop a code for it, like A, B, C, D for example, that the monkey has to remember.

**John Maunsell:** We've done that experiment. The way we did it is not to put the stimulus outside the receptive field, but to take it out to another sensory modality. So in a series of experiments, we had the animal do the task where the sample was never given visually, but through a bar that was mounted onto the front of the chair that he was sitting in. At the start of each trial the computer would rotate that bar to some randomly selected orientation. The monkey would have to grab it, and feel its orientation, since he couldn't see it because there's a plate blocking the view.

Then when he saw a grating with the same orientation come up on the screen, he'd have to release the bar. We recorded from cells, first when they got the visual sample, and then when the animal was using only the tactile sample. For 90 cells, 22 of those, roughly 1/4, had statistically significant effects. All but 4 showed exactly the same pattern of activity as a function of sample, regardless of whether the animal was using a tactile input, or a visual input. What that meant to us is that these cells were encoding information about the orientation the animal had to respond to, and it didn't matter how the animal got this information. It had been extracted beyond the particular sensory modality.

## **What and Where in the Human Brain**

Leslie G. Ungerleider

Laboratory of Neuropsychology, NIMH, NIH, Bethesda, MD 20892

To delineate object and spatial vision pathways in the human brain, we measured regional cerebral blood flow (rCBF) with Positron Emission Tomography (PET) in normal subjects performing object vision (face matching), spatial vision (location matching), and sensorimotor control tasks. The results indicated that the areas with rCBF increases associated with face and location matching had extensive overlap in lateral occipital regions but differed, as expected, in their ventral and dorsal extensions. Whereas cortex with rCBF increases associated with face matching extended anteriorly into the ventral temporal cortex, cortex with rCBF increases associated with location matching extended anteriorly into inferior and superior parietal regions. Thus, in humans, as in monkeys, there appears to be a divergence of visual processing pathways. To determine how the object and spatial information carried separately by the occipitotemporal and occipitoparietal pathways are integrated to yield a unified visual percept, we investigated possible sites of interaction. The results from our neuroanatomical studies in monkeys indicate that zones within the rostral superior temporal sulcus may be sites for convergence of information from the temporal and parietal cortex.

**Ralph Siegel:** What factors are important to find changes in blood flow at these higher order areas?

**Leslie Ungerleider:** To record from these higher order areas, you may really have to drive the system. But we find that effort has a big effect on blood flow.

**Ralph Siegel:** Do you think there may be some correlation with attentional effects?

**Leslie Ungerleider:** I think it's important to note that the subject is able to attend to faces, and ignore locations, or attend to locations and ignore faces, because embedded in this same experiment are conditions which are only a face, or only a location, and we get exactly the same results. So the subject is obviously able to filter out the irrelevant information.

**Peter Lennie:** You showed a picture where you superimposed the PET activity on the MRI scan, and it lit up the fusiform gyrus on the right side of the brain. Was it a consequence of how you presented the stimulus, or does it say something about lateralization?

**Leslie Ungerleider:** That slide may be slightly misleading, because that slide is a subtraction of the activation of face matching minus location matching. So everything in common dropped out except face matching. Plus, we set the threshold to 30% so that it's a focus of activation, it's a very stringent threshold. If one were to lower the threshold, the activation in the left hemisphere would be obvious. It's just that we raised the threshold significantly so that the noise dropped out. But it is the case that the right hemisphere usually is more active than the left, and the activation is more extensive than on the left.

## **Spatial Attention and Spatial Constancy in Posterior Parietal Cortex**

Carol L. Colby

National Eye Institute, NIH, Bethesda, MD 20892, clc@lsr.nei.nih.gov

Visual responses of neurons in posterior parietal cortex are modulated both by overt movements of the eyes and by covert shifts of attention. We have found that these two different kinds of modulation contribute to different cognitive functions. Response modulation by attentional state permits enhanced processing of images within the focus of attention. In contrast, response modulation by intended eye movements makes it possible to maintain perceived spatial constancy of the visual world as images are displaced on the retina. Two neural mechanisms contribute to spatial constancy. First, parietal neurons respond to the memory trace of a visual stimulus when an eye movement brings the spatial location of the stimulus into the receptive field. This memory response indicates that the parietal representation of the visual world is shifted in conjunction with eye movements. Second, some parietal neurons accomplish this shift in anticipation of the actual eye movement. This anticipatory shift may reflect an attentional shift that normally precedes eye movements. An attentional shift alone, however, cannot produce a change in the stored representation. Only when an eye movement is about to occur do we see evidence for a shifted representation. These results suggest that while eye movements and attention normally coincide, the underlying neural mechanisms are distinct and subserve different cognitive functions.

**Ralph Siegel:** You used the words parietal representation and cortical remapping many times. It implies that there is a mapping onto the cortical surface of receptive fields, or what does it imply about the cortical representation onto the surface itself? We have not been able to see any particular type of groupings.

**Carl Colby:** I have also been unable to find crisp and clear maps as there are at earlier stages in the visual system, and when I say remapping, I don't mean not that you're going to see a simple shift as you could see in V2. I don't know how it does work, but it's not going to be a simple shifter circuit.

**Ted Cohn:** Your data indicate that there is a change in spontaneous activity levels depending on the task.

**Carol Colby:** There are changes in activity levels. In fact, when the monkey is in a block of trials and is being asked to make the same saccade over and over, you can begin to see activity, even preceding the onset of the stimulus. There are certainly changes in the baseline activity, when the animal is in a particular behavioral state, as seen in John Maunsell's study with Vincent Ferrara that he just presented.

**Jeff Teeters:** Have you compared the receptive fields of neurons when the stimulus is within the classic receptive field, compared to the receptive field map when you have to go from memory to shift attention.

**Carol Colby:** I do not have a quantitative answer to that question. But, I have noted that in neurons that have a nice tonic activity to this delayed saccade task, you see a nice visual burst and a motor burst, and then this activity in between, and that's when you're on the hot spot, the best spot of the receptive field. As you start moving away from the best spot, it's like the histogram starts sinking below sea level. First you get less tonic activity, good visual and motor bursts, and as you get far enough away you get only a visual burst and a motor burst, and the tonic activity has completely gone away, as though the memory field of the cell is smaller than either the visual field or the saccade field.

## Addendum

The following papers were announced for publication in this proceedings but have been withdrawn or are unavailable.

- [2054-02] **Low level representations in the cortex**  
J. G. Robson, Cambridge Univ. (UK)
- [2054-07] **Distribution of extraretinal neuronal representations across cortical visual pathways**  
J. Maunsell, Baylor College of Medicine
- [2054-09] **What and where in the brain**  
L. G. Ungerleider, National Institutes of Health
- [2054-11] **Neurological mechanisms for localization**  
I. Bodis-Wollner, Univ. of Nebraska Medical Ctr.
- [2054-12] **Structured motion in inferior parietal lobule**  
R. Siegel, Rutgers Univ.
- [2054-14] **Segregation of global vs. local motion processing in primate visual area MT**  
R. T. Born, Harvard Medical School
- [2054-15] **Parietal cortex and spatial constancy**  
C. L. Colby, National Institutes of Health
- [2054-17] **The rotation problem in visually guided navigation**  
M. S. Banks, Univ. of California/Berkeley
- [2054-19] **Effects of neurological damage on complex motion analysis**  
L. M. Vaina, Boston Univ. and Harvard-MIT
- [2054-20] **A neural mode for 2D motion perception and motion transparency**  
H. R. Wilson, Univ. of Chicago
- [2054-31] **Functional organization of color processing in primate visual cortex**  
D. T'so, Baylor College of Medicine
- [2054-32] **The problem of identifying chromatic pathways**  
P. Lennie, Univ. of Rochester

## Author Index

Albrecht, Duane G., 12  
Albright, Thomas D., 211  
Baker, Curtis L., Jr., 104, 124  
Banks, Martin S., Addendum  
Barghout, Lauren, 52  
Bodis-Wollner, Ivan, Addendum  
Bonds, A. B., 1  
Born, Richard T., Addendum  
Boulton, Jane C., 104, 124  
Boynton, Geoffrey M., 32  
Colby, Carol L., Addendum  
Cormack, Lawrence K., 79  
De Valois, Karen K., 95  
De Valois, Russell L., 89  
Derrington, Andrew M., 134  
Foley, John M., 32  
Freeman, Ralph D., 43  
Geisler, Wilson S., 12  
Klein, Stanley A., 142  
Kontsevich, Leonid L., 52, 69  
Kontsevich, Maxim L., 69  
Lawton, Teri B., 194  
Lennie, Peter, Addendum  
Levitt, J. B., 174  
Lund, Jennifer S., 174  
Maunsell, John, Addendum  
Ohzawa, Izumi, 43  
Oram, Michael W., 155  
Perrett, David I., 155  
Robson, John G., Addendum  
Roy, Jean-Pierre, 166  
Schor, Clifton M., 79  
Siegel, Ralph, Addendum  
Stevenson, Scott B., 79  
Stoner, Gene R., 211  
T'so, Dan, Addendum  
Tyler, Christopher W., 52  
Ungerleider, Leslie G., Addendum  
Vaina, Lucia M., Addendum  
Wilson, Hugh R., Addendum  
Wu, Quanfeng, 174  
Young, Malcolm P., 185